A Multidimensional Technique for Sound Quality Assessment

Patrick Susini, Stephen McAdams*, Suzanne Winsberg
Institut de Recherche et de Coordination Acoustique/Musique (TRCAM), 1 place Igor-Stravinsky, 75004 Paris, France

Summary
One approach to improving sound quality is to create a preference map on the basis of several acoustic parameters relevant to auditory perception. The map is derived from several stages of subjective testing, acoustical analysis,
This approach is valid for other types of sound objects. This article proposes a general method for characterizing sound quality on the basis of objective criteria. The different stages of the method as well as the precautions to be taken in using it will be discussed below, with a study performed on interior car sounds being used as an example [11, 12]. This study was performed at IRCAM in collaboration with the automobile manufacturers Renault and PSA Citroën and falls under the framework of a contractual agreement. Due to confidentiality constraints, the description will remain formal without explicitly describing the results obtained (identity of sound sources and exact nature of acoustic parameters). This constraint does not, however, affect the main aim of the article, which is the presentation of the experimental methods and data analysis techniques employed to explore sound quality in all its natural sonic complexity.

A global view of the method showing the relations among the different stages is presented in Figure 1. At the outset, the sound corpus must be defined and the sounds carefully selected in order to obtain a relatively homogeneous corpus (section 2). As a function of the class of sound object(s) being studied, the panel of listeners must be selected according to appropriate criteria (section 3). Globally, a first step consists in determining the perceptual attributes common to the panel of listeners that are used to compare sounds to one another (section 4). A multidimensional scaling analysis (CLASCAL) yields a spatial model that can be represented graphically and which reveals the perceptual dimensions (section 5). In the last stage, the degree of preference (or, inversely, annoyance) associated with each sound is evaluated as a function of the perceptually significant acoustic parameters revealed in the previous stage (section 6). The advantage of this approach is that it does not limit the exploration and characterization of the components of sound quality to parameters that are already known. It provides a method for finding new perceptually salient parameters that engineers had not imagined, but that listeners hear and use in their evaluations anyway.

2. Establishment of the Sound Corpus

In order to ensure a realistic restitution of the sound field in the experimental situation that is as close as possible to the original conditions, it is necessary to use binaural techniques [13]. The car sounds used in the study were recorded digitally by the automobile manufacturers with a dummy-head system seated in each car in the front passenger seat with the car running on a test track. The sounds were reproduced in the experiments by IRCAM's ISPW real-time digital signal processing environment [14], using the MAX application on the NeXT workstation. All sounds were edited to 5 s in duration with 50 ms attack and decay ramps, taking care to avoid sounds with recording artifacts. The three important points for sound restitution are the following:

- the specifications of the binaural recording conditions
Table I. Composition of four groups of car sounds used.

<table>
<thead>
<tr>
<th>Loudness variable</th>
<th>Motor speed 1</th>
<th>Motor speed 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loudness variable</td>
<td>Group 1</td>
<td>Group 2</td>
</tr>
<tr>
<td>Loudness equalized</td>
<td>Group 3</td>
<td>Group 4</td>
</tr>
</tbody>
</table>

According to the type of study, it is generally advisable to determine which categories best represent the judgments of subjects in the main experiments. For this purpose, category scaling was used with the CLASCAL program. However, in cases where a single rating for a parameter is not sufficient, multidimensional scaling analysis can be used. The CLASCAL program is designed for this purpose and is used in the second part of the study to analyze the loudness, pitch, and broad-band parameters of the sounds.

4. Dissimilarity Scaling: The CLASCAL Program

4.1. Introduction
stimulus i and stimulus j developed by Winsberg and De Soete [9, 17] and presented in Figure 1, postulates common dimensions shared by all stimuli, specific attributes or specificities particular to each stimulus, and latent classes each of which have different saliences or weights for each of the common dimensions and the set of specificities. Maximum likelihood estimates of the model parameters are determined (that is the coordinates of each stimulus on each dimension of the common space, the specificities of the stimuli, the dimension weights for each class, and the proportion of subjects in each class). An EM (expectation-maximization) algorithm is used to determine the class structure. (See [9] for a complete description of the CLASCAL algorithm and the latent class approach to MDS. Space does not permit an adequate description here.) The class structure is latent; there is no apriori assumption concerning the latent class to which a given subject belongs. Model selection, including the choice of the appropriate number of latent classes, the number of common dimensions and the presence or absence of specificities, is based on the BIC statistic which is information based and depends on the log likelihood, the number of model parameters, and the number of observations [18], as well as on a Monte Carlo procedure [19]. Models with the lowest BIC values are chosen. The Monte Carlo procedure is used to determine the appropriate number of latent classes and to verify the spatial model. (See [9] for a discussion on model selection.) The CLASCAL analyses yield a spatial representation of the N stimuli on the R dimensions, the specificities of each stimulus, the probability that each subject belongs to each latent class (generally equal to one for one of the classes) and the weights or saliences of each perceptual dimension for each class. Also included in the output results are the model variance, the log likelihood, the value of the BIC statistic, and the results of the Monte Carlo procedure if used.

4.4. Results

A single class was sufficient for all data sets. An analysis with the number of dimensions varying between 1 and 6 without specificities and between 1 and 5 with specificities was then performed for a single subject class. BIC indicated that the best model was three-dimensional with specificities for Group 3 and two-dimensional with specificities for the other groups. A graphic representation of the data corresponding to the sounds of Group 3 is given in Figure 2. Each sample is represented in the three-dimensional common space. The specificities of the stimuli are not shown in the figure.

5. Physical Parameters Underlying the Perceptual Space

5.1. Acoustic analysis
Table II. Correlation coefficients between perceptual coordinates and "objective" parameters. The probability $p$ that the two measures are independent is indicated for $p < 0.01$ (*) and for $p < 0.05$ (†). Parameter $\varphi_1 = \text{Loudness}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varphi_1$</td>
<td>Dim. 1</td>
<td>Dim. 2</td>
</tr>
<tr>
<td>$\varphi_2$</td>
<td>Dim. 1</td>
<td>Dim. 2</td>
</tr>
<tr>
<td>$\varphi_3$</td>
<td>Dim. 1</td>
<td>Dim. 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varphi_2$</td>
<td>Dim. 1</td>
<td>Dim. 2</td>
</tr>
<tr>
<td>$\varphi_3$</td>
<td>Dim. 1</td>
<td>Dim. 2</td>
</tr>
<tr>
<td>$\varphi_4$</td>
<td>Dim. 1</td>
<td>Dim. 2</td>
</tr>
</tbody>
</table>

rectangular bandwidth) formula was adopted for auditory filter bandwidths and Patterson et al.'s [23] gammatone filters were used. On the bases of these early analysis stages, stimulus set. The preference analysis to be examined in the next section will determine the utility for preference of each of the predictor parameters. Those that have a strong weight can thus in turn be equalized over the sound set in a subsequent stage of experimentation. As such, this approach allows a progressive minimization of perceived differences in terms of preference or unpleasantness between the stimuli.


6.1. Procedure

At the beginning of the session, the subject listened to all the samples in a random order to get a sense of the range of variation possible. Then, all pairs of different samples were presented in a random order. On each trial the subject heard a pair of samples only once and had to choose which sound was preferred. The data consist of the proportion of subjects that preferred one sample over another for each of the N stimuli.
The functions $f$ and $g$ are splines, that is, piecewise polynomial functions, the number and order of each polynomial being variable. The order of the spline is taken to be equal to the maximum degree of the polynomial plus one. The functions between the polynomials of a given spline correspond to the "internal knots" and are of finite number. Attention is restricted to splines with maximal continuity at each knot. That is, for a spline of order $k$, the function and its $k - 2$ derivatives are continuous. The default knot positions are located at a ladder of quantiles, thus an equal number of observations is used to determine each spline parameter. Each spline of a given degree can be represented as a linear combination of a set of basis functions.

In the multivariate model (3), the utility is a function of both parameters:

$$u = f(a, b).$$

This multivariate function is represented by a multivariate spline, that can be defined as a linear combination of tensor products of univariate basis splines. In this model therefore, the effect of one parameter on utility depends on the value of the other parameter.

For a preference matrix and a given set of predictor parameters, a large number of models may be tested by varying the type of model (additive or multivariate) and the order and number of internal knots of the splines. BIC is used to choose between the candidate models. For a given preference matrix, this criterion can also be used to choose from among various pairs of predictor parameters that had more or less similar correlations with the perceptual space. This latter choice is of a heuristic nature, since the correlations with the different perceptual dimensions are also taken into account. In general, the best model is chosen of parameters in terms of their interpretability. The analyses yield the preference function for each dimension for an additive model, and the preference surface for a multivariate model.

As an example, the utility functions for acoustic level and another parameter (parameter 2) are shown in Figure 3 in which the additive model was retained. Not surprisingly, the utility function for level is more or less monotonic: the more the level increases, the more the utility for preference decreases. The variation in utility with the other parameter is nonmonotonic, however its overall influence on preference is comparatively small and is completely overpowered by the level variation. For the level-equated sample sets, the parameters had more equivalent influence on preference and some of these were clearly nonmonotonic, indicating that there are zones in the middle of the range of variation that are maximally or minimally preferred by the panel of listeners.

7. Conclusions

The data analysis reveals representations of the sets of car sounds that have several perceptual dimensions for both combinations of gear setting and motor speed. The loudness equalization allowed the emergence of other factors contributing towards similarity perception and quite notably to preference. When loudness is a factor in the comparison of car sounds, it clearly dominates the preference judgments. The role of the secondary parameters vary with gear setting and motor speed and globally have less importance in these judgments. Generally, the perceptual salience of the dimensions is not the same for the two gear setting/motor speed combinations, indicating that the relative importance of perceptual cues evolves with changes in car functioning and context (equalized loudness or not).

This method, when applied to very different sets of sound objects (musical instruments and car sounds), thus reveals: 1) a multidimensional mental representation of complex sound events, 2) a close relation between the perceptual dimensions and acoustic properties or their auditory transforms, and 3) specific properties of certain sounds that affect their perception and comparison with other sounds. For musical instrument sounds, differences between listeners in the perceptual importance according to the different dimensions and specificities have also been found. While such differences were not found in the studies on car sounds mentioned above, they should be systematically verified in such experiments in case they do exist. The coherence of the results across these different stimulus sets indicates a great stability in the perceptual processes involved in the appreciation of homogeneous sets of sounds. Further, it is important to emphasize that aside from loudness, none of the objective parameters revealed corresponded to those currently included by default in most sound quality measuring devices. A cautionary note on a methodological issue is warranted here how-
structure was found that was strongly categorical in nature and which was associated with a high degree of identification of the individual sound sources by the listeners [26]. Judgments of dissimilarity seemed in this case to be based entirely on the categories that the stimuli belonged to and less to their individual acoustic and sensory properties. Thus, this predominant cognitive factor—recognition, classification, and identification of the sound source [12]—rendered inappropriate the conception of the perceptual classification, and identification of the sound source [12]—rendered inappropriate the conception of the perceptual space in terms of continuous underlying dimensions. The above techniques of preference analysis are not applicable in such cases and other techniques need to be used. However, for homogenous sound sets, the experimental approach described above is a powerful method for understanding the perceptual underpinnings of sound quality.

References