Modelling the relationship between human perception and Sound Quality parameters using LS-SVMs

T. Coen, N. Jans, P. Van de Ponseele, I. Goethals, J. De Baerdemaeker, B. De Moor
K.U.Leuven, Department of Electrotechnical Engineering,
Kasteelpark Arenberg, B-3001, Heverlee, Belgium
e-mail: **tom.coen@agr.kuleuven.ac.be**

Abstract
The increasing pressure on the design cycle of an automobile makes the classic solution of jury testing no longer acceptable. A model of the human perception of engine sounds allows faster and more frequent feedback.
In this paper the relationship between judge background and judge scores, as well as between car characteristics and judge scores is examined.
Subsequently a model to classify cars on comfortability and sportiness based on the Sound Quality parameters of their engine sound is developed. Finally a model to compare two cars on comfortability and sportiness is drawn up.
Comfortability can be modelled accurately. Lack of a suitable Sound Quality parameter renders modelling sportiness very hard.

1 Introduction

In recent years the relationship between automobile manufacturers and consumers has changed tremendously. The design of a car has become more and more based upon the desires of the consumer. Since consumer desires are subject to change over time, the design specifications of a car change as well. This necessitates shorter design cycles in order to keep up with customer desires [1] [2].

In this paper the focus lies upon the engine sound and the perception of this sound by the potential consumer. In order to obtain the opinion of the consumer, jury tests have to be organized. In such a test, a person is asked to score each sound on a characteristic, for example comfortability.

There are several drawbacks to the classic practice of jury testing, which make it incompatible with the current evolution of the automobile industry:

Disturbances: Variation of equipment, different interpretation of the questions, noise,. . . introduce a judge-specific bias to the scores.

Composition of the jury: For a significant jury test a large and balanced (different background, age, . . .) population is needed.

The above mentioned problems result in a considerable time span (about a month) that is needed to organize and process a jury test [3] [4]. This is no longer deemed acceptable.

Objective In this paper a model will be developed that predicts human perception of engine sounds based upon Sound Quality (SQ) parameters of those sounds. Nine SQ parameters that are generally accepted as relevant in the automobile industry are selected as input for the model.

LS-SVM was chosen as modelling technique.

The selected SQ parameters are A-weighted Sound Pressure Level (SPLA), B-weighted Sound Pressure Level (SPLB), Zwicker Loudness, Articulation Index (AI), Modified or Open Articulation Index (AIM), ANSI Speech Interference Level (ASIL), Preferred Speech Interference Level (PSIL), Sharpness and Roughness.

Different outputs are defined. A model can pass a quantitative (for example a grade between 0 and 10) or a qualitative (different classes of cars, for example good, normal and bad) judgement. A model can also compare two cars. In this paper a model for qualitative judgement of cars (Section 4) and for comparison of two cars (Section 5) will be presented.

2 Data acquisition and exploration

2.1 Experiments

Run-ups of 30 significantly different cars were recorded doing road tests [5]. The sound was recorded to the left and right of the head support of the driver. In this way the recorded sound is the actual sound heard in the car by the driver. This set-up implies that the engine sound as well as the effect of the isolation of the interior of the car is taken into account. Note that the opinion of the driver is considered important here.

The recorded sounds are then used in a jury test. Two characteristics of engine sounds will be examined: comfortability and sportiness. The participants fill out a form with some background information (age, driving habits,'car perception',. . .) and grade each sound twice. The sound is played and the participants then give a grade between 0 and 10 on both characteristics. Each sound is graded twice to check the consistency of the judge. If those two grades are too far apart on too many cars on either one of the questions (comfortability or sportiness), the scores of this judge are removed from the dataset.

The jury test consists of 104 judges. The dataset used here is based on the average score given by the 79 judges that are consistent on both characteristics.

2.2 SQ parameters

2.2.1 Definition

The different SQ parameters can be divided in three groups.

A first group of parameters, namely SPLA, SPLB and Zwicker Loudness, is correlated with the Sound Pressure Level of the sound. SPLA and SPLB are Sound Pressure Levels with respective weighing functions A and B [6]. Zwicker Loudness is the human perception of sound, and is calculated from SPL levels by using a conversion table [7].

The second group, namely AI, AIM, ASIL and PSIL [8], describes how comprehensible a conversation would be with the sound as background. AI and AIM are based on a special weighing of the SPL levels. Frequencies that are more important for the understanding of speech receive a higher weighing factor. The results are normalized. 100% means that a conversation is perfectly comprehensible. 70% or less means that conversation becomes difficult.

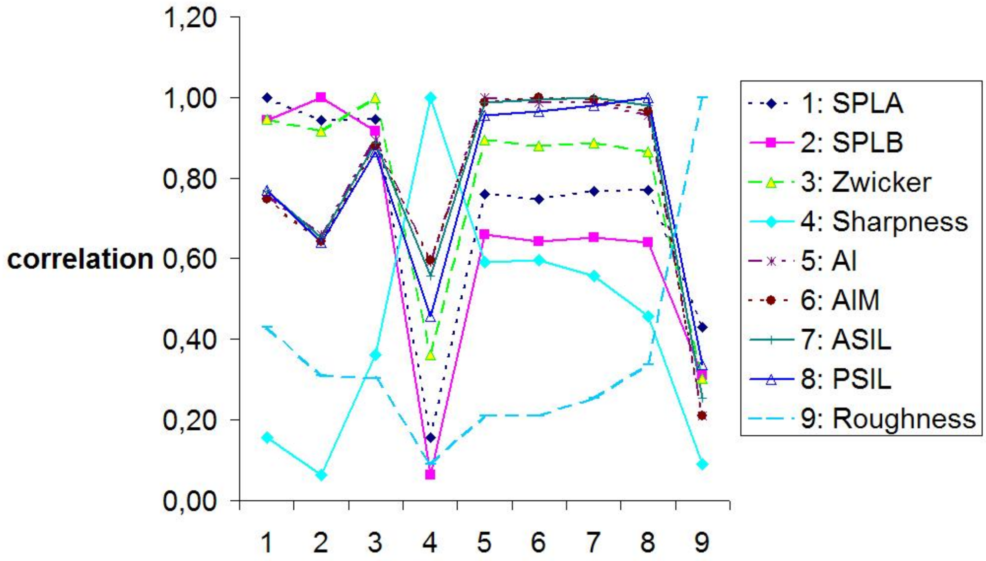

Figure 1: Normalized absolute mutual correlation for each of the SQ parameters

ASIL and PSIL are the average of the SPL levels over the frequency bands that are important for speech. Thus, the lower the value of this parameter, the more comprehensible a conversation is.

It is clear from the definition that there is a negative correlation between the AI, AIM and ASIL, PSIL.

A third group of parameters consists of Sharpness and Roughness [8]. Sharpness is based on the Loudness algorithm with higher weighting factors for the higher frequencies. Roughness is a measure of the degree of modulation weighted per third octave of the sound.

It is clear that not all these parameters are independent. Within the first and second group there is a strong correlation between the defined parameters. In a later stage of the modelling the most appropriate parameters will be selected.

In Figure 1 the normalized absolute correlation between each of the SQ parameters for all the measured cars is shown. The correlation within the first and second group of parameters is illustrated.

2.2.2 Relevance

In this section the relevance of the different parameters for the prediction of the human perception of a sound is examined. This can be done in two ways.

Correlation with scores The normalized correlation with the scores (averaged over all judges) of each of the parameters is shown in Table 1.

It is clear from these results that sportiness scores will be far more difficult to predict (based on these SQ parameters) than comfortability scores. For comfortability scores especially the parameters from the first and second group are significantly correlated with the scores. The above mentioned negative correlation between AI, AIM and ASIL, PSIL is also visible in these results.

	comfortability	sportiness
SPLA	-0.95167	0.39788
SPLB	-0.91952	0.30018
Zwicker	-0.94031	0.34893
Sharpness	-0.15915	0.17922
AI	0.78781	-0.40565
AIM	0.78017	-0.40505
ASIL	-0.81080	0.41977
PSIL	-0.81891	0.40840
Roughness	-0.33459	0.34147

Table 1: Normalized correlation between SQ parameters and comfortability scores or sportiness scores

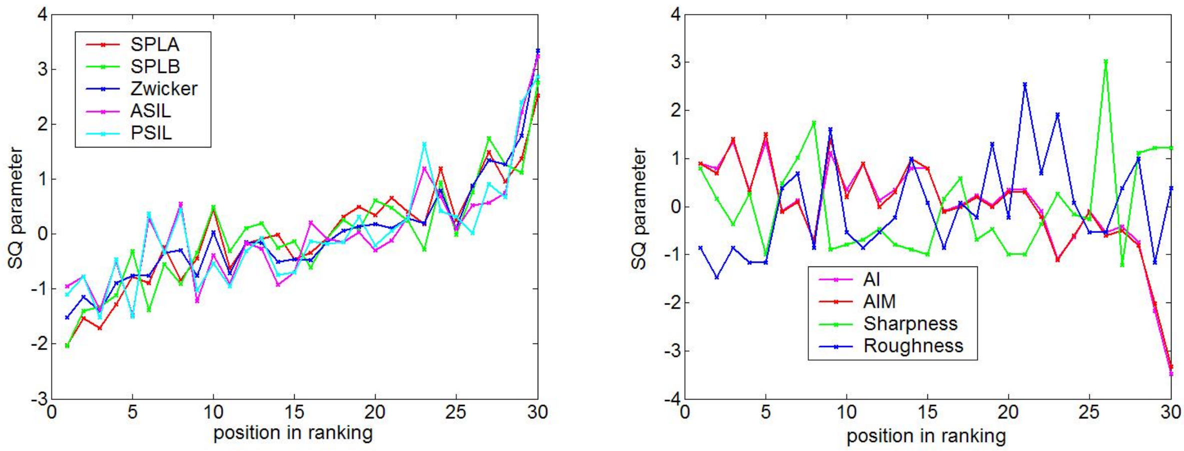

Figure 2: Evolution of the normalized SQ parameters over the comfortability ranking

Ranking Based on the grades given by the judges, a ranking of the cars for comfortability and sportiness can be established. In Figure 2 the nine SQ parameters are plotted in function of the position of the car in the ranking. The scores of the cars decrease from left to right in the plot.

For comfortability there is a clear trend. All parameters that are correlated with the sound level in the car (SPLA, SPLB, Zwicker Loudness) are at a minimum for the most comfortable car. As indicated by the correlation between SQ parameters and comfortability scores (see above), the SQ parameters of the first and second group exhibit the clearest relationship with the ranking. There is no correlation between the ranking and the parameters of the third group (Sharpness, Roughness).

For sportiness there is no clear trend visible. This is mostly due to lack of a good parameter for sportiness. This sportiness parameter is still the topic of ongoing research [9]. Often a mechanical parameter, namely rpm, is used to get a parameter for sportiness [4].

Automatic Relevance Determination ARD [10] is used to determine which SQ parameters are the most important for predicting human perception of an engine sound. ARD is a special form of Least Squares Support Vector Machines (LS-SVMs) with Radial Basis Function (RBF) kernel (see Section 3.1) which enables weighing the different inputs. The weight assigned to an input is proportional to its importance. For comfortability Zwicker Loudness, ASIL, AIM and SPLB are the most important parameters. For sportiness the algorithm confirms the lack of a good parameter. All parameters are equally (un)important.

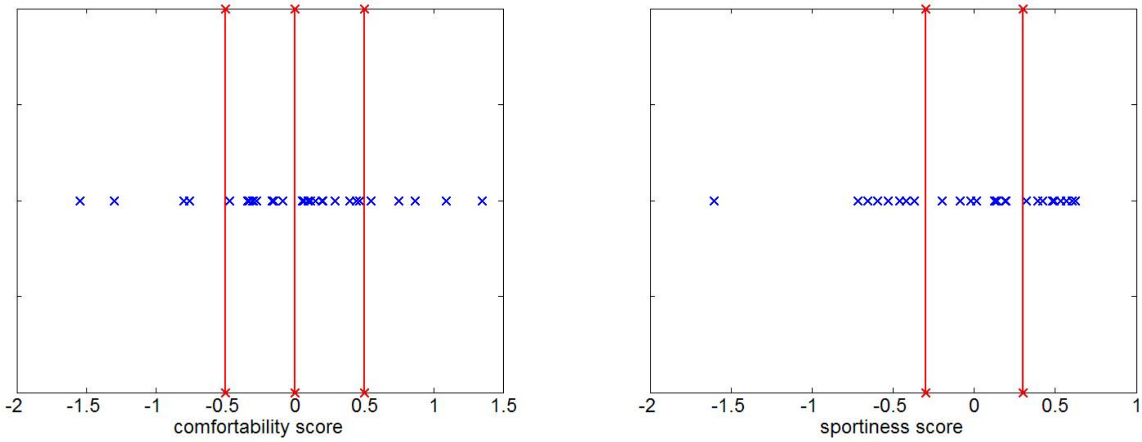

Figure 3: Normalized scores of the 30 cars on comfortability (left) and sportiness (right)

2.3 Scores

2.3.1 Normalization

The judges have given scores on comfortability and sportiness between 0 and 10. These scores are normalized to zero mean and unit standard deviation and then averaged over all the judges. This is done because the judges are no car experts, and thus the variation and mean of their scores (over all the cars) is not significant. A judge with a high variation would have a greater impact on the final score of a car (relative to the other cars).

The normalized scores for comfortability and sportiness are shown in Figure 3. Notice the clusters in the scores. A possible definition of classes is indicated by vertical lines. These are the thresholds used for the classifiers of Section 4. There are not enough datapoints to determine whether the visible structure is real or coincidal. This structure will however influence the performance of models.

2.3.2 Background of the judge

It has been reported that jury tests performed by experts (sound engineers of car manufacturers, ...), show a significant influence of the background of a judge (age, education, ...) on the scores accorded by the judge [11].

For the significant population of the here described jury test however, no relationship between background and accorded scores could be found. This was examined for different characteristics of judges such as: age, gender, perception of a car, driving experience, education, ...

The comfortability and sportiness scores of all 79 consistent judges are plotted for several cars and labelled with the judge characteristic. If there is some relation between the judge characteristics and the accorded scores, clusters should be visible for at least some of the cars. This is not the case for any of the cars in the test.

The averages of each group of judges is also plotted. These averages are always close together for the different groups. An example of these plots is given in Figure 4 for different ‘car perceptions’. This also illustrates the enormous variation over the different judges.

Based on these jury tests, it can be concluded that for a general population, there is no (clear) relationship between judge background and judge scores.

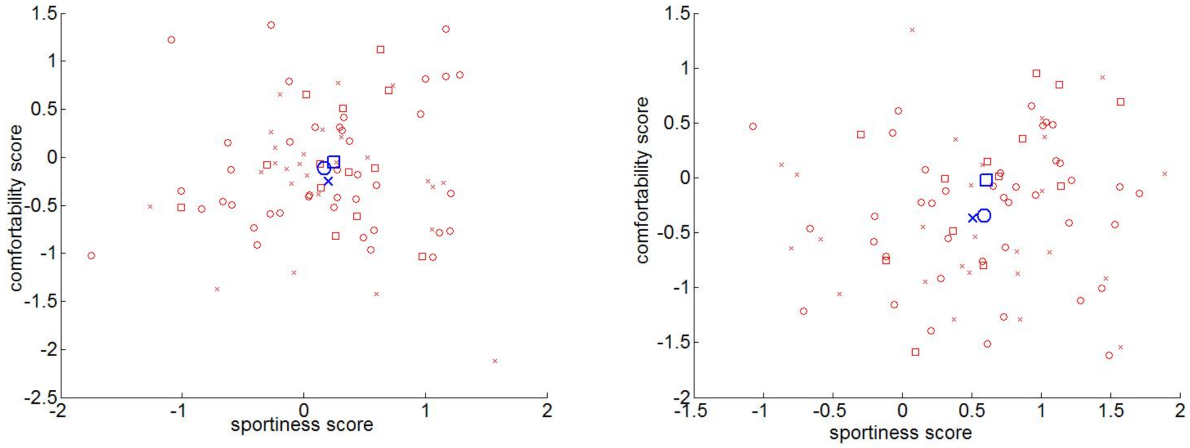

Figure 4: Labelling of comfortability (on Y) vs sportiness scores (on X) plot with 'car perception' by a judge, for two different cars ('x' = A way to get from point A to B, '○' = An easy and comfortable way to travel, '□' = An extension of your personality, large symbol indicates the average of the group)

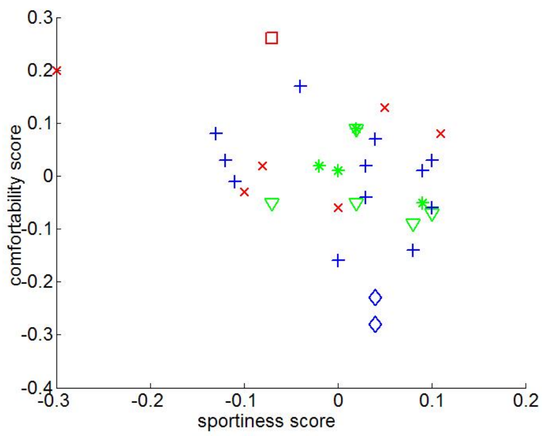

Figure 5: Comfort scores (on Y) and sportiness scores (on X) for all cars labelled with the car characteristics ('x' = sedan / '+' = break / '□' = SUV / '◊' = transporter / '∗' = small car / '▽' = monovolume)

2.3.3 Car characteristics

The relationship between the type of car (break, sedan, transporter, ...) and the score on comfortability and sportiness was also examined. In Figure 5 the comfortability and sportiness scores of all 30 cars are shown labelled with the type of car.

There is no clear pattern visible. Only the transporters are clearly recognized by the judges. This is to be expected since the engine of these cars is hardly isolated.

3 Modelling

For each of the 30 cars there are three datavectors available:

- SQ vector: the values of the nine calculated SQ parameters

- comfortability scores: the normalized scores on comfortability (zero mean and unit standard deviation) of the 79 consistent judges
- sportiness scores: the normalized scores on sportiness (zero mean and unit standard deviation) of the 79 consistent judges

Separate models will be defined for the prediction of comfortability and sportiness using the SQ vector as input for the models.

3.1 Modelling technique

Least Squares Support Vector Machines (LS-SVMs) is selected as modelling technique. This is a neural networks technique that can be used for classification as well as for function estimation. LS-SVM for function estimation can be derived from classification with minor adjustments.

As an illustration, a classification problem with two classes is assumed. Classes are labelled with -1 and 1. LS-SVMs for other classification problems or for function estimation are very analogous. For more information see [10].

LS-SVM defines a hyperplane with a weight vector w and a bias b. Given a dataset x_k, y_k with $k = 1..N$, x_k being an input vector and y_k being the class to which this vector belongs, this hyperplane has to satisfy following border conditions:

$$\begin{aligned} y_k[w^T x_k + b] &= 1 - e_k, \qquad k = 1..N, \\ e_k &= \text{classification error on point } k. \end{aligned} \tag{1}$$

To obtain a good classifier, the number of misclassifications needs to be minimized. This leads to the following optimization problem:

$$\min_{w,b} \mathcal{J}(w, b, e),$$

with

$$\begin{aligned} \mathcal{J}(w, b, e) &= \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2, \\ y_k[w^T x_k + b] &= 1 - e_k, \qquad k = 1, \ldots, N. \end{aligned} \tag{2}$$

$w^T w$ is a regularization term to avoid overfitting and γ is the regularization constant. The dividing hyperplane then is $w^T x + b = 0$. The hyperplane is defined in such a way that as many points as possible of class 1 lie on the straight line $w^T x + b = 1$ and of class -1 on $w^T x + b = -1$.

The above model is linear. Using the Mercer condition [12] this theory can be extended to non-linear models. The input data is then transformed by a transformation φ to an higher dimensional input space (possibly even infinitely dimensional) where the classes are linearly separable. This extension leads to a new set of border conditions, namely:

$$y_k[w^T \varphi(x_k) + b] = 1 - e_k, \qquad k = 1..N. \tag{3}$$

With the method of Lagrange multipliers this can be transformed into an optimization problem without constraints:

$$\max_{\alpha} \min_{w,b,e} \mathcal{L}(w, b, e; \alpha),$$

linear kernel	$x_k^T x$	(Lin)
polynomial kernel	$(x_k^T x + 1)^d$	(Poly$d$)
RBF kernel	$e^{\frac{-\|x - x_k\|_2^2}{\sigma^2}}$	(RBF)

Table 2: The different kernel functions and their abbreviation in this paper

with

$$\mathcal{L}(w, b, e; \alpha) = J(w, b, e) - \sum_{k=1}^{N} \alpha_k \{y_k [w^T \varphi(x_k) + b] - 1 + e_k\}. \tag{4}$$

This leads to a system of linear equations. An explicit construction of the above used transformation is not needed, and the value of the kernel $K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l)$ suffices.

The expression for the classifier now is:

$$y(x) = \text{sign}[\sum_{k=1}^{N} \alpha_k K(x, x_k) + b]. \tag{5}$$

The weighing factor α_k is called the support value of datapoint k. There are several possibilities for the kernel function K. In this paper the linear, polynomial and RBF kernel will be used (see Table 2).

3.2 Model design decisions

A number of parameters have to be chosen for a classification model. There are the number of classes and the threshold between the different classes.

A kernel function has to be selected. The parameters of this kernel is automatically determined (by means of a non-convex optimization).

A final important parameter is the encoding used for the target classes. LS-SVMLab offers several encodings such as MOC (Minimum Output Encoding), OnevsOne (OO) and OnevsAll (OA) [13]. Instead of such an encoding, function estimation (FE) can be used as modelling technique followed by a discretization in classes. Both approaches were used.

3.3 Repeatability

Training of LS-SVMs given regularization constant γ and kernelparameters, is a convex problem with a unique solution. This solution however is dependent on the used dataset. Since the available data is divided in a trainingset (which is used for modelling) and a testset (which is used to estimate the performance of a model), the composition of the trainingset has to be randomized. For each experiment 10 runs with different randomly generated composition are performed. In this paper the trainingset always consists of 75% of the cars.

In each run the regularization constant (γ) and the kernel parameters are determined through a non-convex optimization. This optimization is based on a crossvalidation procedure.

The median of the correct classification percentages (on the testset) over the 10 runs gives the best indication of the model performance (because it is less sensitive to outliers than for example the mean). The median will be used in the rest of this paper to compare model performance.

	MOC	MOC4	OA	OA4	FE	FE4
Lin	71.4%	78.6%	92.9%	78.6%	85.7%	92.9%
RBF	85.7%	92.9%	50%	42.9%	100%	100%
Poly2	64.3%	78.6%	0%	7.1%	85.7%	92.9%
Poly3	57.1%	71.4%	0%	0%	71.4%	85.7%

Table 3: Median for different kernels and encodings, two classes comfortability

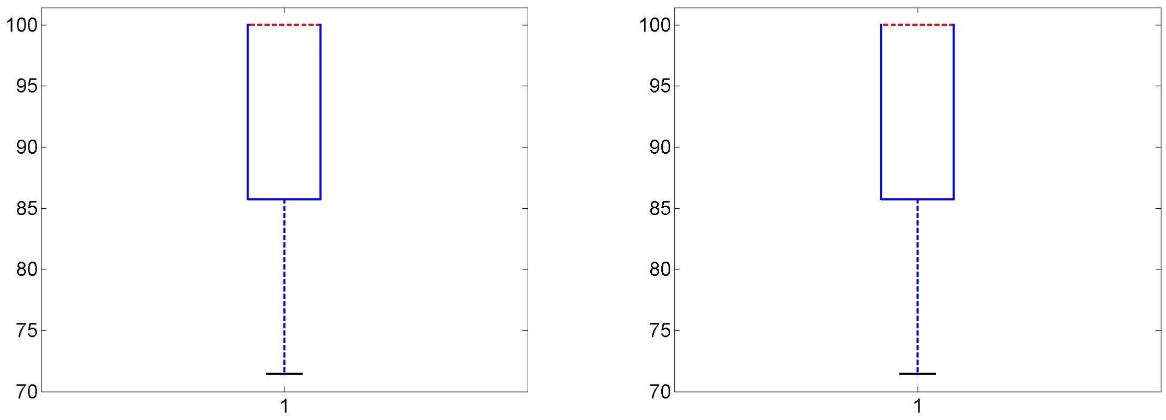

Figure 6: Boxplot for classification comfortability scores in two classes with FE , RBF kernel and nine inputs (left) or four inputs (right). Both models appear to be equivalent.

4 Qualitative judgement

4.1 Comfortability judgement

4.1.1 Classification in two classes

Two classes are defined as score larger than and smaller than 0. The number of datapoints for each class is:

	$s < 0$	$s > 0$
Number	14	16

The medians for the performed experiments are shown in Table 3. The best results are obtained with RBF kernel and FE. The linear kernel with OA encoding also leads to very good results.

Using the results obtained with ARD (see Section 2.2.2), this classifier can be modelled based on the four most important inputs (which are Zwicker Loudness, SPLB, AIM and ASIL). These results are also shown in Table 3 (marked with 4 in the heading). Reducing the input space results in a significant improvement of the results for FE. The best results are obtained with FE and RBF kernel, but also linear and polynomial kernel (order 2) perform well.

The boxplots for FE and RBF kernel with nine respectively four inputs are shown in Figure 6. They are quasi identical for both experiments. Notice the small interquartile range and large first quartile (85.7%) which indicates a good model performance.

	MOC	OO	OOb	OA	OAb	FE
Lin	42.9%	28.6%	57.1%	42.9%	57.1%	71.4%
RBF	42.9%	0%	57.1%	21.4%	42.9%	71.4%
Poly2	28.6%	0%		0%		42.9%
Poly3	21.4%	0%		28.6%		57.1%

Table 4: Median for different kernels and encodings, three classes comfortability

	Lin MOC	Lin FE	RBF MOC	RBF FE
4 inputs	57.1%	71.4%	50%	85.7%
9 inputs	42.9%	71.4%	42.9%	71.4%

Table 5: Comparing the medians for nine and four inputs, three classes comfortability

4.1.2 Classification in three classes

Three classes are defined as followed: scores clearly larger than 0 (larger than 0.25), scores clearly smaller than 0 (smaller than -0.25) and scores around 0 (between -0.25 and 0.25). The number of datapoints in each class is:

	$s < -0.25$	$-0.25 < s < 0.25$	$s > 0.25$
Number	8	13	9

Experiments with different kernels and encodings are performed. The medians of the experiments are shown in Table 4.

Only FE with linear kernel and RBF kernel gives acceptable results. The bad results come from the lack of robustness of the OO and OA encodings. If one bit goes wrong during the prediction, there are two equivalent possibilities for the decoder. Instead of choosing one, the decoder returns infinity. The results improve by adapting the Hamming decoder to take one of the possibilities. These experiments are marked with a b (OOb and OAb).

Reducing the dimension of the input space to four by only including SPLB, Zwicker, ASIL and AIM in the input vector, leads to better results. The medians of these experiments are shown in Table 5.

FE with RBF kernel now reaches a median of 85.7%. The boxplots for FE with RBF kernel with nine and four inputs are compared in Figure 7. The interquartile range is analogous for both experiments. The first quartile is however significantly larger for the model with four inputs.

4.1.3 Classification in four classes

Four classes are defined: scores greater than 0.5, scores between 0 and 0.5, scores between -0.5 and 0 and scores less than -0.5. The number of datavectors in each class is:

	$s < -0.5$	$-0.5 \leq s < 0$	$0 \leq s < 0.5$	$s \geq 0.5$
Number	5	11	10	4

Experiments with linear kernel and RBF kernel are performed. The medians for the different kernels and encodings are shown in Table 6.

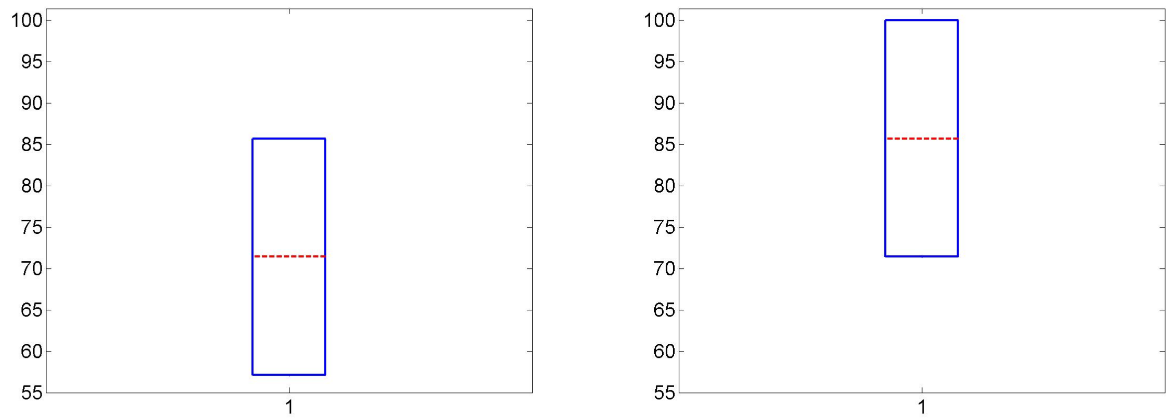

Figure 7: Boxplot for classifying comfortability scores in three classes with FE, RBF kernel and nine inputs (left) or four inputs (SPLB, Zwicker, ASIL and AIM) (right) The interquartile range does not change, the first quartile is however significantly larger with four inputs.

	MOC	OO	FE	FE4
Lin	35.7%	50%	78.6%	85.7%
RBF	28.6%	28.6%	85.7%	92.9%

Table 6: Median for different kernels and encodings, four classes comfortability

FE gives clearly the best results. FE is also used with only the four most important inputs (FE4). The results improve significantly.

As mentioned in Section 2.3.1, there are clusters visible in the comfortability scores of Figure 3. These clusters coincide with the four classes defined here. This explains the improved performance of the classifier with four classes relative to the classifier with three classes.

The boxplots for FE with four inputs and respectively linear and RBF kernel are shown in Figure 8. The median is larger for the RBF kernel case than for the linear kernel case. The interquartile range is smaller for the RBF kernel as well.

4.2 Sportiness judgement

4.2.1 Classification in two classes

Two classes are defined. Scores larger than 0 and scores smaller than 0. The number of datapoints in each class is:

	$s < 0$	$s > 0$
Number	10	20

Experiments are performed with linear, RBF and polynomial kernel combined with MOC encoding, OA encoding and FE. The medians are shown in Table 7. The best results are now obtained with RBF kernel and MOC encoding. FE gives poor results because of the increased complexity of the relationship between sportiness and SQ parameters.

The boxplots for linear and RBF kernel with MOC encoding are shown in Figure 9. The median of the linear kernel coincides with the first quartile, whilst the first quartile of the RBF kernel is much smaller

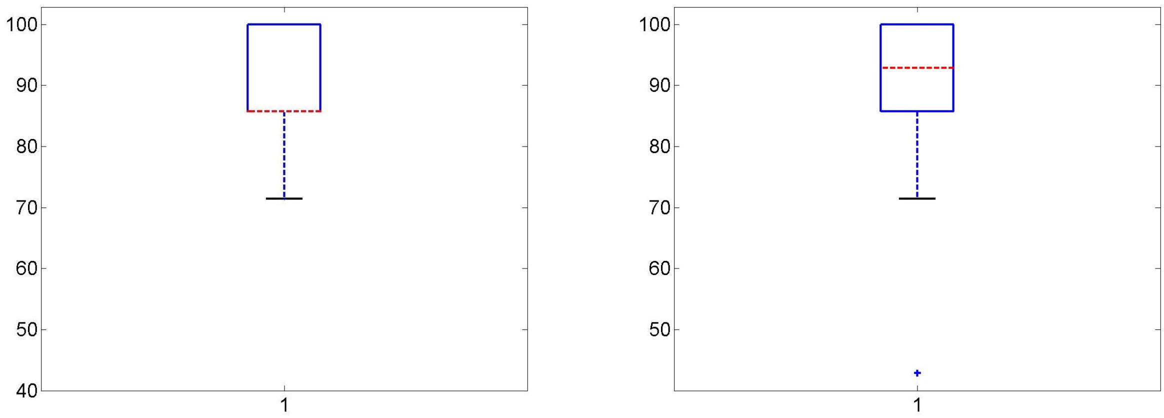

Figure 8: Boxplot for classifying comfortability scores in four classes with FE, four inputs and linear kernel (left) or RBF kernel (right). With RBF kernel the mediaan is larger and the interquartile range is smaller.

	MOC	OA	FE
Lin	71.4%	42.9%	50%
RBF	76.6%	57.1%	64.3%
Poly2	42.9%	0%	57.1%
Poly3	42.9%	0%	42.9%

Table 7: Median for different kernels and encodings, two classes sportiness

than the median. Despite the slightly larger median of the RBF kernel, the linear model has a better overall performance.

4.2.2 Classification in three classes

Three classes are defined as in Figure 3. Scores larger than 0.3, scores between -0.3 and 0.3 and scores smaller than -0.3. The number of datapoints in each class is:

	$s < -0.3$	$-0.3 \leq s < 0.3$	$s \geq 0.3$
Number	8	13	9

The medians obtained for different kernels and encodings are shown in Table 8. The special version of the Hamming function was used everywhere.

It is clear that the results are poor everywhere. Because of the higher complexity of the relationship between SQ parameters and sportiness, not even FE can provide help here.

5 Comparing two cars

The SQ vectors of two cars are used as input for the model. The output is a relative judgement of the sportiness or comfortability of both cars. This kind of model can be used to establish a ranking of cars, and to fit a new car into an existing ranking.

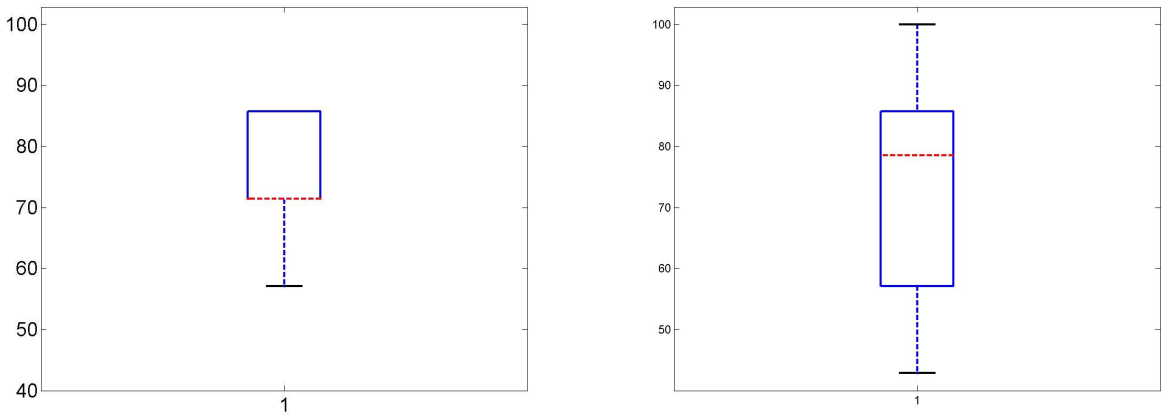

Figure 9: Boxplot for classifying sportiness scores in two classes with MOC encoding and linear kernel (left) or RBF kernel (right) The first quartile of the linear kernel coincides with the median, which makes the linear kernel the best option

	MOC	OO	OA	FE
Lin	28.6%	28.6%	28.6%	35.7%
RBF	21.4%	42.9%	28.6%	35.7%
Poly2	28.6%	42.9%	14.3%	28.6%
Poly3	28.6%	35.7%	28.6%	21.4%

Table 8: Median for different kernels and encodings, three classes sportiness

The dataset was divided in two groups for every run: a trainingset of 23 cars and a testset of 7 cars. Within each group every car is compared to all other cars in order to compile the actual dataset. The trainingset thus contains $\frac{(22+1)22}{2} = 253$ datavectors, the testset contains $\frac{(6+1)6}{2} = 21$ datavectors.

Models are trained with two different input configurations: two SQ vectors as input (dimension 18) and the difference of both SQ vectors as input (dimension 9) (indicated by Δ).

5.1 Comparing two cars on comfortability

Experiments are performed with RBF and linear kernel, combined with MOC encoding and FE (followed by discretization). The medians of the here discussed experiments are shown in Table 9.

The results improve significantly by applying the difference of both SQ vectors to the input. There are two reasons for this:

- The dimension of the input space is halved from eighteen to nine.
- The composition of the input is more obvious. In the case with eighteen inputs the corresponding SQ parameters of both the cars have to be mapped to one an other. This adds complexity to the modelling task.

Since the difference of SQ vectors is a good input, the change in scores must be quasi linear with the change in SQ parameters.

	MOC	FE	ΔMOC	ΔFE
Lin	82.1%	82.1%	85.7%	89.3%
RBF	82.1%	67.9%	85.7%	89.3%

Table 9: Median for comparing cars on comfortability

	MOC	FE	ΔMOC	ΔFE
Lin	60.7%	60.7%	71.4%	71.4%
RBF	53.6%	64.3%	64.3%	64.3%

Table 10: Median for comparing cars on sportiness

5.2 Comparing two cars on sportiness

The results are shown in Table 10.

The results are clearly less good here than in the comfortability case. This is because of the higher complexity of the relation between SQ parameters and sportiness scores. An extra SQ parameter with a better correlation to sportiness scores would improve the model significantly.

The difference of SQ vectors as input gives again slightly better results. This implies that sportiness scores are quite linear with SQ parameters as well.

In Figure 10 the ROC-curves are shown for comparing cars on comfortability and sportiness, both modelled with linear kernel and ΔMOC. The area under the ROC-curve is clearly larger for the comfortability case. This illustrates again the enormous difference in performance of both models.

6 Conclusion

The examined judge characteristics have no clear influence on the scoring of judges for a general population. There is no correlation found between car characteristics and judge scores (except for a transporter).

Using LS-SVM to classify and compare cars on comfortability gives good results. Lack of a suitable SQ parameter complicates the modelling of sportiness. Comparing of cars on sportiness however, delivers reasonable results.

The visible structure of the judge data (Figure 3) helps to choose the right thresholds for the classes. This is illustrated by the better performance of the 4 classes case relative to the 3 classes case for comfortability.

Comparing cars enables the ranking of cars. This approach is also more robust. Comparing cars on sportiness gives better results than classifying cars on sportiness.

It is obvious that a new SQ parameter with a high correlation with sportiness is needed. An example of such a parameter is rpm extraction of engine sound [9].

The models can be made more robust by using more data.

Acknowledgements

The research work was done while the senior author was at K.U.Leuven, Department of Electrotechnical Engineering (ESAT). The research was conducted in cooperation with LMS International (**http://www.lmsintl.com**).

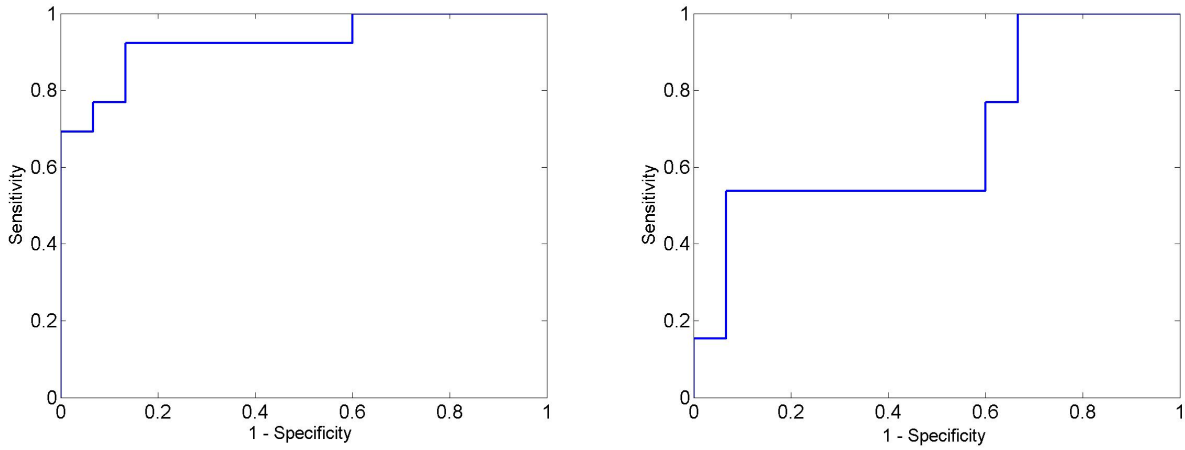

Figure 10: ROC curve for comparing cars on comfortability (left) and sportiness scores (right) with ΔMOC and linear kernel. The area under the curve shows that the comfortability model is clearly better.

References

[1] P. Schöggl, *Confort acoustique et vibratoire automobile et ferroviaire*, Société des Ingénieurs Automobile, Courbevoie 3-4 December (1998).

[2] W. Keiper, *Sound quality evaluation in the product cycle*, Acustica, Vol. 83, No. 5, pp. 784-788 (1997).

[3] H. Fastle, *The psychoacoustics of sound-quality evaluation*, Acustica, Vol. 83, No. 5, pp. 754-764 (1997).

[4] R. Genuit, *Background and practical examples of sound design*, Acustica, Vol. 83, No. 5, pp. 805-812 (1997).

[5] T. Coen, N. Jans, I. Goethals, P. Van de Ponseele, A. Vecchio, B. De Moor, *Modelleren van het verband tussen menselijke waarneming en Sound Quality parameters gebruik makende van LS-SVMs*, Katholieke Universiteit Leuven, Departement Elektrotechniek, Leuven (2004).

[6] H. Van der Auweraer, K. Wykaert, *Sound quality: perception analysis and engineering*, Internal report, LMS International, Leuven.

[7] E. Zwicker, *Procedure for calculating loudness of temporally variable sounds*, Journal of the Acoustical Society of America, Vol. 62, No. 3, pp. 675-682 (1977).

[8] H. Van der Auweraer, P. Van de Ponseele, *Sound quality evaluation and measurement criteria*, Internal report, LMS International, Leuven.

[9] I. Belghit, A. Vecchio, K. Janssens, *Plausibility check of rpm measurements*, *Proceedings of the Eleventh International Congress on Sound and Vibration, St. Petersburg, Russia, 2004 July*, St. Petersburg (2004).

[10] J.A.K. Suykens, T. Van Gestel, B. De Moor, J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific Pub. Co., Singapore (2002).

[11] N.C. Otto, B.J. Feng, G.H. Wakefield, *Sound Quality Research at Ford - Past, Present and Future*, Sound and Vibration, 20-24 (1998).

[12] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, Germany, Berlin (1995).

[13] http://www.esat.kuleuven.ac.be/sista/lssvmlab/