Subjective and Objective Sound Quality Predictive Models for the Assessment of a Propeller Aircraft Interior Noise

Claudio Colangeli; Bernardo Lopes; Agnieszka Mroz; Karl Janssens and Herman Van der Auweraer[1]

[1] *Siemens Industry Software NV, 3001 Leuven (Belgium), E-Mail:claudio.colangeli@siemens.com*

Abstract

The interior sound of an aircraft is generally not optimized for passenger acoustic comfort because its assessment often occurs in the late stages of the development cycle. To improve these aspects, the design philosophy should shift to a human-centered paradigm and the study of the sound quality aspects front-loaded. In this paper we discuss a data-driven method for the evaluation of a propeller aircraft interior noise on the basis of objective and subjective psychoacoustic attributes. Such tool, combined with virtual prototyping and sound synthesis tools, paves the way for the inclusion of the human perception in the aircraft design optimization process. The developed instrument grounds on a modular approach capable of classifying in terms of objective sound quality attributes and of subjective passenger annoyance the inputted propeller aircraft in-cabin sound samples, obtained through experimental recordings or through sound synthesis of numerically simulated data. The objective sound quality features are estimated through a set of Convolutional Neural Network models trained on time domain labeled data, while the subjective annoyance is predicted through a feature-based Artificial Neural Network, trained on the basis of a jury test campaign. The paper discusses the accuracy of the method and reports on experimental applications.

Introduction

In this paper we describe a modular approach for computing the salient interior Sound Quality attributes of a propeller aircraft adopting data-driven models (**Figure *1***).

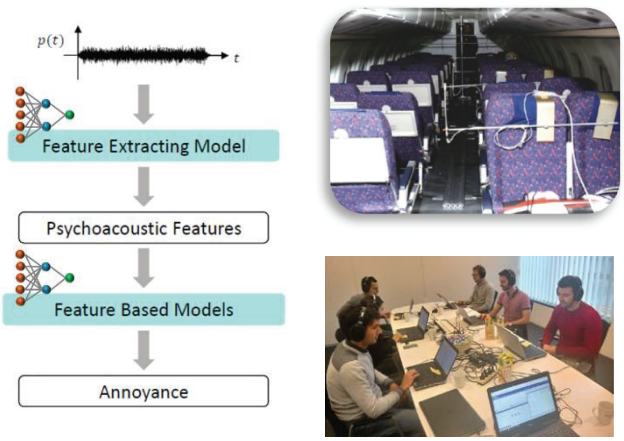

Figure 1: the purpose of this study is the development of data-driven models for the prediction of subjective and objective sound quality attributes on the basis of experimental recordings and jury test evaluations.

The predictive model consists of two modules. The first module adopts Convolutional Neural Networks (CNNs), for the prediction of psychoacoustic metrics, directly from pressure signals. The second module adopts an Artificial Neural Network for the prediction of passenger acoustic discomfort, i.e. *Annoyance*, when exposed to a given propeller aircraft interior noise. This second module relies on objective sound quality attributes as inputs and returns, as output, the estimation of a passenger *Annoyance*, computed on the basis of experimental data preliminarly obtained through a subjective evaluation test. Once trained, the obtained data-driven model allows the prediction of the passenger *Annoyance* directly from raw time data, by-passing the extraction of intermediate features. This is an advantage in case the model is adopted in multi-attribute optimization loops [1] or embedded in control logics [2].

The proposed modular approach is hereafter described and its performance assessed on the experimental example of a propeller regional aircraft interior noise

Sound Quality analysis

The analyzed experimental data were recorded at 5000 ft altitude, during the cruising phase of the flight. The studied aircraft is equiped with two 4-blades propellers (more details in reference [3]). Interior sound was measured – in the frame of a previous study – in 85 in-cabin positions at passenger ear level (circa 1.2 m above the floor) in two different propeller settings: *Synchronous* and *Asynchronous*. In the first case the rotational speed of the two propeller was synchronized, in the second case it was not. In this latter case, therefore, the correlation between the noise sources generated by the two propeller is reduced and the fundamental and blade-passing frequencies slightly shifted, influencing the perception attributes of the consequent in-cabin noise.

For each sound sample, five psychoacoustic metrics were computed: *Specific Loudness* (ISO 532B Diffuse Field, first chritical band considered: 0-100 Hz) *Loudness* (ISO 532B Diffuse Field), *Sharpness* (DIN54692 Diffuse Field), *Tonality* and *Roughness* [4]. A spatial mapping of each one of these features can be seen in **Figure *2***. The major expected in-cabin noise contributors are the propellers tonal noise and the broadband sound induced by the TBL-related excitation. The computed sound quality attributes reveal that the front part of the aircraft closer to the propellers is mainly influenced by their tonal noise. In this case, therefore, the dominant attributes are *Loudness* and *Tonality*. The TBL broadband contribution influences mostly the rear part of the cabin – where its contribution dominates over the propeller tonal noise – and in the very frontal section of the cabin, where TBL noise is dominant below 100 Hz. This results, in the first case, in an increase of the *Roughness* metric and in the second case in an increase of the *Bark 1* metric. The *Sharpness* metric reveals an increase of the high frequency range in the external regions of the rear section of the cabin.

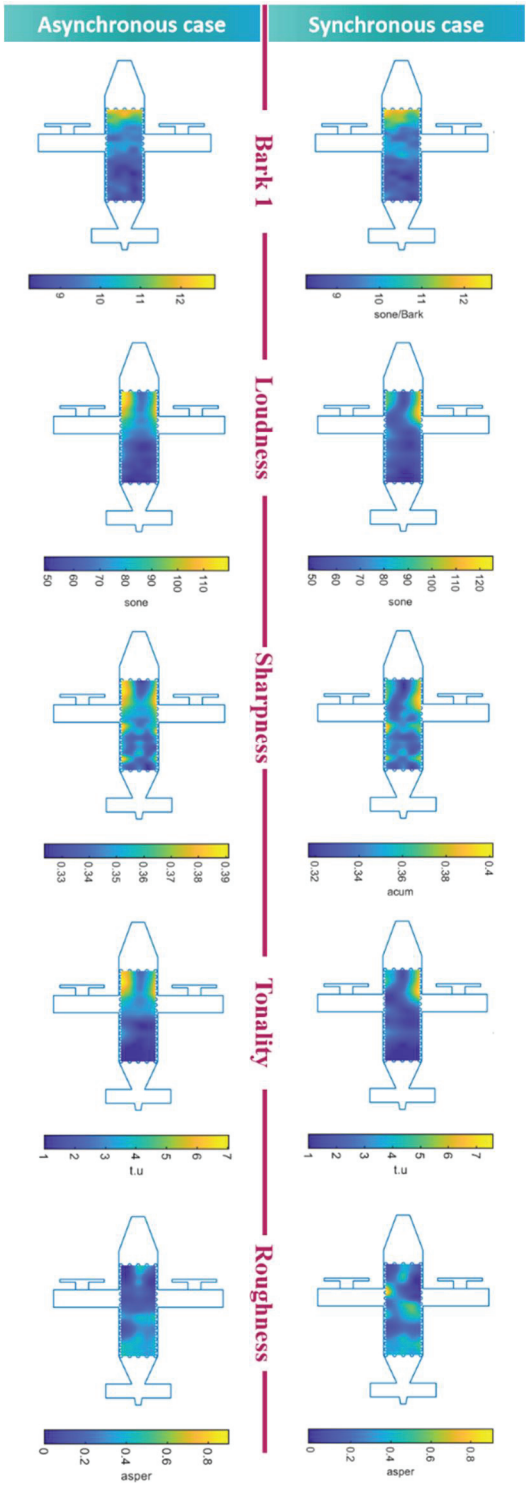

Figure 2: the computed sound quality attributes on the recorded in-cabin sounds in the *syhcronous* and *asynchronous* configurations.

A *k-means* clustering analysis (ref. [5]) of the computed psychoacoustic features was performed in order to identify homogeneous regions within the aircraft cabin. **Figure *3*** reports the results adopting four clusters. The obtained clusters clearly identify two distingushed regions: one close to the propellers and one in the rear of the aircraft. The remaining two clusters identify two more distributed regions.

The performed sound quality analysis enabled to extract and map the computed objective psychoacoustic features along the aircraft cabin. The result of the clustering analysis, moreover, enabled the identification of a criterion for the selection of a sub-set of in-cabin sounds to be adopted for performing a subjective evaluation of the *Annoyance* perceived by the passenger.

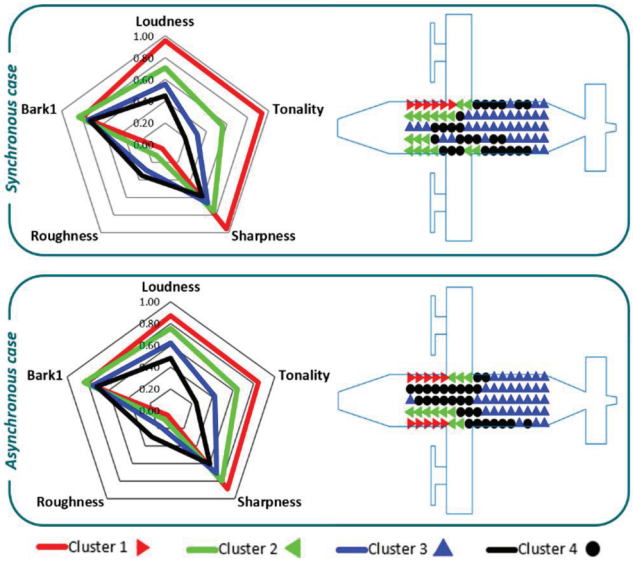

Figure 3: the result of the *k-means* clustering analysis performed on the five metrics reported in the spider graphs, identifies four distinguished areas within the aircraft cabin.

Assessment of subjective annoyance

We assessed the subjective perception of *Annoyance* of the passenger in the studied aircraft through a jury test based on the *anchor semantic differential* approach [6]. In this case the juror is asked to evaluate the current sound sample with respect to a reference sound (*anchor*) on a scale going from *Much Less Annoying* to *Much More Annoying* (**Figure *4***). In order to maximize the homogeneity of the quality of the sounds replayed in the jury test, the recorded sound samples were synthesized (see reference [3] for more information on the adopted synthesis technique). The chosen *anchor* sound was synthetically generated in order to obtain average sound quality attributes (features similar to the case of cluster 3 in **Figure *3***). 30 sounds were selected for the jury test: 2 every cluster and 22 showing prominent behavior of singular features. The consistency of juror evaluations was checked by including 5 repeated sound samples (guidelines from reference [7] were followed). The evaluated sound samples had 6.5s duration (including fade-in and fade-out of 50 ms) and have been replayed after the *anchor* sound with a pause of 1.5s in between. A total of 40 jurors were involved in sessions of not more than 6 jurors at a time.

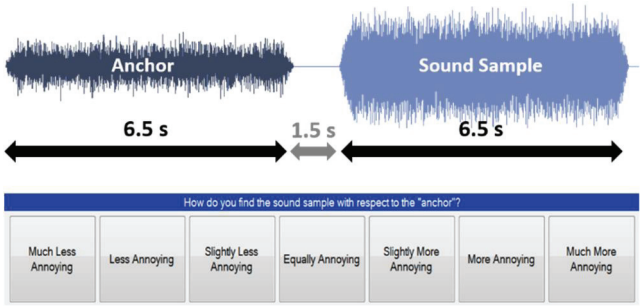

Figure 4: the jury test was carried out with 40 jurors and 30 sounds, adopting the *anchor semantic differential* approach.

Each juror classified the sound stimuli with respect to the *anchor* sound by choosing one of the available adjectives (**Figure 4**). The results were standardized according to the *standard deviation* of each juror and then, for each stimulus the average evaluation from all jurors was computed. Finally, each stimulus evaluation was re-scaled in a way that 100 corresponds to maximum *Annoyance* value and 0 to its minimum value (**Figure 5**). This procedure was also used, in the following steps related to the data-driven modeling, for the psychoacoustic metrics whose correlation with Annoyance is reported in ***Table 1***.

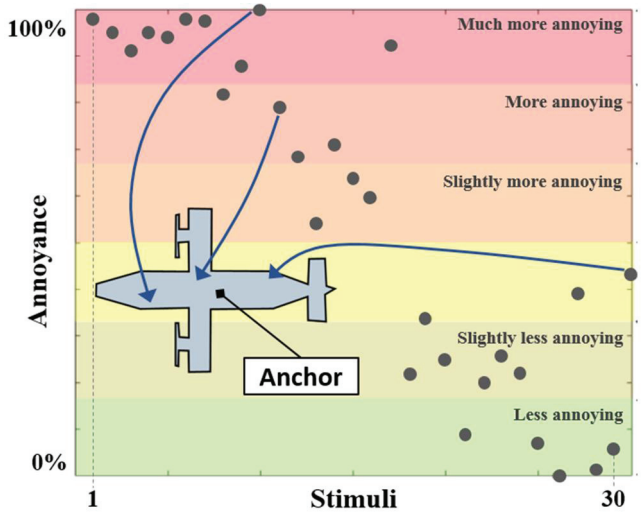

Figure 5: the result (mean value of the 40 subjective evaluations) of the jury test on the 30 sounds analyzed, has normalized on a continuous scale from 0% to 100% before using the data in the machine learning models. The figure reports the position within the cabin of four salient sounds.

Table 1: correlation between the adopted objective and subjective sound quality metrics.

	N	S	T	R	A
LOUDNESS (N)	1	0.99	0.98	-0.66	0.93
SHARPNESS (S)	-	1	0.97	-0.62	0.94
TONALITY (T)	-	-	1	-0.72	0.87
ROUGHNESS (R)	-	-	-	1	-0.53
ANNOYANCE (A)	-	-	-	-	1

Feature-based predictive model

Even though there is a high correlation between some psychoacoustic metrics and *Annoyance*, the use of one single metric does not allow to model acoustic discomfort. A combination of features is necessary for building a model able to correlate psychoacoustic features with results from jury studies. In order to do so, we performed a comparative study between 4 different prediction techniques: Multiple Linear Regression (MLR) [8]; Support Vector Machines (SVMs) [9]; the combination of several decision trees as a Random Forest (RF) and Artificial Neural Networks (ANN) [5]. Numerous training techniques can be used. The Bayesian Regularization [10] and the Levenberg-Marquardt (LM) algorithm [11] were compared in this study and finally the LM was adopted. On the basis of a sensitivity analysis performed adopting the LM training technique, the data were divided into 70% for training and 30% for testing. The performance of the four feature-based machine learning appraoches (MLR, SVM, RF, ANN) was tested through a Monte Carlo simulation. The *Coefficient of Determination* (R^2) and the *Rooth Mean Square Error* (RMSE) have been adopted as performance metrics of each prediction model obtained with 100 random data divisions (70% training data, 30% test data). The results of this analysis show that ANN outperforms the other methods (***Table 2***). For this reason this approach was further adopted for the creation of the sought modular predictive model.

Table 2: averaged performance metrics of the four tested feature-based approaches.

	R^2	RMSE
MLR	0.862	15.323
SVM	0.92733	8.637
RF	0.972	6.216
ANN	0.98185	5.018

The comparative study revealed, moreover, that the best performing ANN architecture is the one adopting 2 neurons in the hidden layer of the network. **Figure 6** reports the results of the best performing ANN predictive model obtained with such an optimized architecture, inputing the full set of computed Sound Quality metrics.

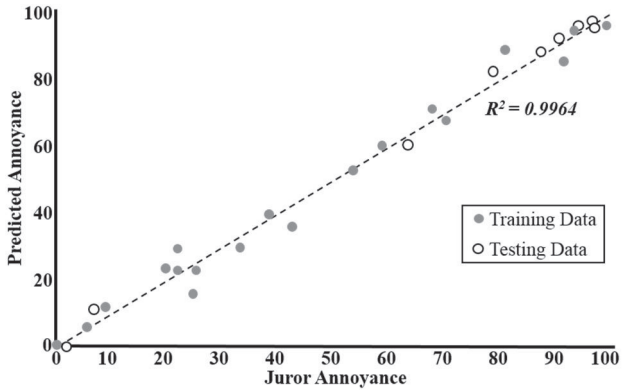

Figure 6: the best performing ANN predictive model ensures a *Coefficient of Determination* of $R^2 = 0.9964$.

Feature-free predictive models

Feature-free predictive models like CNNs allow the prediction of a designated output on the basis of raw data such as time recording or images [12]. These approaches require in general a training database larger than ANNs. This is a severe limitation when modelling subjective attributes. Also in the presented case, in fact, the 30 jury test data avaiable were not sufficient for the training of a CNN capable of modelling the passenger *Annoyance* directly from sound recordings. In order to still enable the prediction of the passenger *Annoyance* from time domain inputs, a modular architecture was conceived in which CNNs are adopted for the data-driven computation of the objective psychoacoustic attributes adopted, in the previons section, for the generation of the ANN predictive model. The prediction of attributes such as Loudness, Sharpness, Tonality and Roughness through CNNs is in fact more plausible thanks to the availability of a larger dataset (all the 170 recorded in-cabin sounds). From the global set of 170 sound samples, the 30 sounds used for jury testing were used for assessing prediction performance. The remaining 140 stimuli have been adopted for training. The optimization of the architecture and hyper-parameters of the networks was

done using Bayesian Optimization. ***Table 3*** summarizes the performance of the developed models.

Table 3: performance of the CNN-based models of the objective sound quality attributes in the studied case.

	R^2	*RMSE [%]*
LOUDNESS	0.9094	10.9950
SHARPNESS	0.8863	11.9310
TONALITY	0.9407	10.2450
ROUGHNESS	0.6345	13.8610

The introduction of a feature-free module for the computation of the sought psychoacoustic attributes has on the one hand the advantage of enabling the prediction of *Annoyance* from time signals, on the other hand it introduces an additional source of inaccouracy. This might degrade the overall performance of the modular model. For this reason another Monte Carlo simulation was carried out to determine the best performing architecture. In this analysis, the CNN models have been sequentially removed one-by-one considering all the possible combinations, incuding those with one single psychoacoustic feature. The 30 samples from the jury study were used for performance assessment. For each feature combination, the predicted features are introduced into 100 differently trained ANNs. The overall performance was therefore computed (comparing with the original juror response) and averaged. This study revealed that the best performing modular scheme foresees the feature-free prediction of Loudness and Sharpness and a consequent ANN composed by 2 neurons in the hidden layer (***Figure 7***).

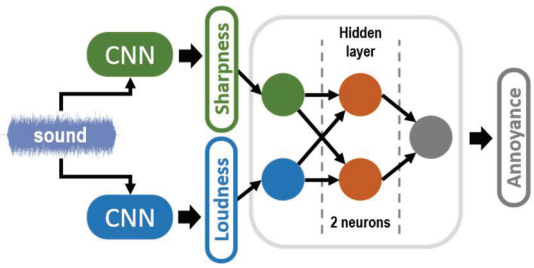

Figure 7: the best performing modular scheme consists of two parallel CNN for the computation of Loudness and Sharpness in series with an ANN (2 neurons in hidden layer).

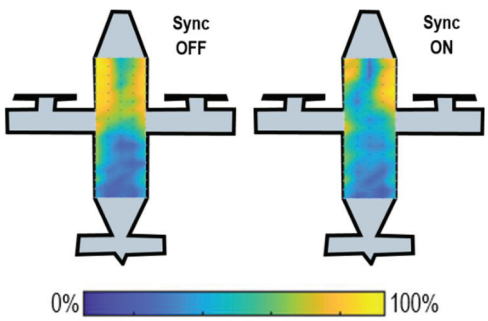

Figure 8: map of the *Annoyance* predicted in the aircraft cabin adopting the computed modular predictive model.

The best performing predictive model obtained through the optimized modular scheme allows a *Coefficient of Determination* of $R^2 = 0.9060$ (***Figure 9***). This is reduced with respect the feature-based case (***Figure 6***), but still satisfactory for a reliable estimation of sought perceived in-cabin *Annoyance* (**Figure *8***).

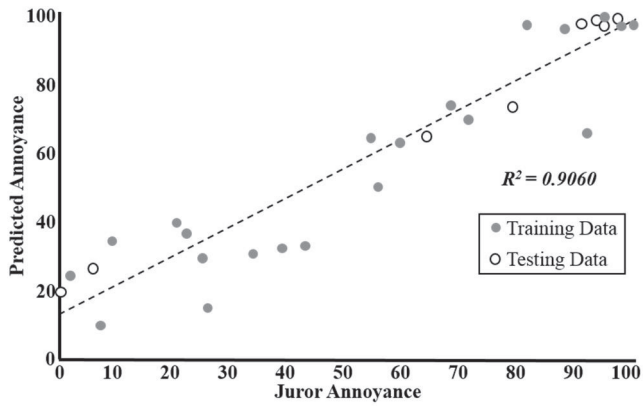

Figure 9: best performing modular data-driven model.

Conclusions

The combined use of feature-free and feature-based data-driven models has proven to be effective in the correlation of propeller aircraft in-cabin sound recordings and their perceived *Annoyance*, obtained on a limited range of cases through subjective evaluation. This enables, in future steps, the easier integration of the developed predictive models into a multi-attribute optimization workflow of a propeller aircraft virtual prototype and on on-board active noise control logics. In this perspective, ongoing studies are devoted to the assessment of the validity of the developed predictive models when applied to different aircrafts of similar characteristics and in similar flight conditions.

References

[1] Lefebvre, T. et al. “Methodological enhancements in MDO process investigated in the AGILE European project.” 18th AIAA/ISSMO Conference. (2017).

[2] Quaranta, V. et al. ‘3D internal noise simulation for vibro-acoustic comfort assessment: an application proposal for AW-109 helicopter.” J. of Aero. Sc. (2011).

[3] Janssens, K., Vecchio, A., and Van der Auweraer, H. “Synthesis and sound quality evaluation of exterior and interior aircraft noise.” Aerosp. Science and Tec. (2008).

[4] Zwicker, E. and Fastl, H. “Psychoacoustics: Facts and models”. Vol. 22. Springer, (2013).

[5] Bishop, C. M. “Pattern recognition and machine learning.” Springer, (2006).

[6] Huang, Hai B., et al. "Sound quality prediction of vehicle interior noise using deep belief networks." Applied Acoustics 113 (2016): 149-161.

[7] Otto, N. et al. “Guidelines for jury evaluations of automotive sounds.” SAE transactions (1999).

[8] Rawlings, J. O., Pantula, S., and Dickey, D. A. “Applied regression analysis: a research tool.” Springer, (2001).

[9] Vapnik, V. “The nature of statistical learning theory.” Springer, (2013).

[10] MacKay, D. J. C. “Bayesian interpolation.” Neural computation 4.3 (1992): 415-447.

[11] Zhang, H. F. et al. “Performance of the Levenberg–Marquardt neural network approach in nuclear mass prediction.” J. of Physics G: Nuclear and Part. Ph (2017).

[12] LeCun, Y. et Al. “Deep learning.” Nature (2015).