inter.noise
HAMBURG 2016

# Perceptual evaluation of the sound quality of car HVAC systems

Antoine MINARD[1]; Christophe LAMBOURG[1]; Patrick BOUSSARD[1]
[1] GENESIS, France

## ABSTRACT

For electric and hybrid vehicles the engine sound has become a less important concern because its level is now much lower. However this reduction made other sound sources clearly audible, such as air conditioning. The CEVAS project deals with the simulation and the perception of the sounds of air conditioning (HVAC) and battery cooling (BTM) systems in cars. In the framework of the project a psychoacoustic study involving a semantic differential listening test was carried on in order to identify the perceptual space of car HVAC sounds and link it to perceived unpleasantness. Twelve semantic scales were presented to 20 participants in order to describe 60 HVAC sounds representative of the existing variety of systems. A principal component analysis identified a 3-dimensional perceptual space explaining more than 95% of the variance in the data. Finally regression analyses on the unpleasantness of the car HVAC sounds revealed the expected strong influence of loudness, and more occasionally tonality and fluctuation strength.

Keywords: Automotive HVAC, sound quality   I-INCE Classification of Subjects Number(s): 63.7,69.3

## 1. INTRODUCTION

The aim of the CEVAS project is to improve the acoustic modeling of automotive air-conditioning systems (HVAC) and to study their perception. The final goal is to provide an efficient tool for the design of new HVAC systems, which are optimized in regard to acoustic comfort criteria that have become more important with the generalization of silent vehicles (either electric of hybrid).

A part of the project is dedicated to the aeroacoustic characterizing and modeling of HVAC systems (1, 2), up to the audio synthesis of the produced sound (3). Another part deals with the sound perception of automotive HVAC. The study detailed here was realized within this specific framework of the project and addresses in particular the relevant auditory attributes for describing HVAC sounds and their relation to the unpleasantness that this type of sound can provoke.

Scientific literature offers a variety of studies dealing with the acoustic comfort that is associated to HVAC sounds, usually in contexts of either dwellings or offices. Some studies (4, 5, 6, 7), based on the use of questionnaires and on-site measurements, only attempt to relate sound annoyance caused by HVAC systems to sound level indicators. Other works rather use laboratory experiments in order to better control potentially influent factors of unpleasantness. Thus several studies on residential HVAC systems (8, 9, 10) revealed a timbre space made of 2 or 3 dimensions, often mainly explained by sound sharpness and tonality. These studies also proposed sound quality models based on the calculation of psychoacoustic indicators that correspond to these percepts. Usually, listeners tend to prefer sounds that are dull or muffled (low sharpness) and not tonal, even if some studies reveal different groups of listeners whose preferences are different, even sometimes opposite.

In the field of transport, the sound of air-conditioning in trains was also studied, by Kahn and Hogstrom (11) among others. These authors also found a link between unpleasantness and both sharpness and tonality. However other studies (12, 13), with a cognitive approach to acoustic comfort inside trains, showed a favorable effect of the HVAC sound for the inner ambiance: this sound allows passengers to feel isolated from other passengers' conversations or unwanted noises. In the automotive context, the topic of HVAC sound was addressed by Menager and Rochepeau (14) who studied 5 automotive HVAC systems, recorded either in vehicle or in an anechoic room, with different modes of operation. The obtained prediction model links unpleasantness to loudness (detrimental to sound quality) and Speech Interference Level (SIL, favorable to sound quality). Finally, the project CESAM also dealt with automotive HVAC in a laboratory context (15). Twelve HVAC systems were considered.

[1] aminard@optis-world.com

They were recorded both on test bench and in vehicle, at 2 operating modes and 2 airflow rates. Several listening tests were conducted that resulted in a pleasantness model based on Speech Interference Level and Aures' model of tonality (16).

In the scope of project CEVAS, the work described here aimed at studying an enlarged set of recordings of automotive HVAC, with a larger number of systems that includes models of hybrid and electric vehicles, and a larger variety of operating modes. This dataset was studied through a comprehensive psychoacoustic methodology from the definition of relevant semantic scales to the design of a robust unpleasantness model, that was eventually compared to the previously mentioned models.

## 2. PERCEPTUAL EVALUATION OF AUTOMOTIVE HVAC SOUNDS

The principle of the present study is to identify the different sound dimensions of automotive HVAC systems. For this, a *semantic differential* experiment (17) was conducted. This particular type of experiment consists in rating the sounds on several scales whose boundaries are defined by pairs of adjectives or nouns of opposite meanings.

However, in order not to create a bias through the use of a predefined vocabulary whose relevance is only arbitrarily defined by the experimenter, the used semantic scales were identified by conducting a preliminary verbalization experiment with the method used by Nosulenko et al. (18). Ten listeners were asked to verbalize the similarity and the preference between each sound pair from a set of 8 recordings. The analysis of the results of this experiment (not detailed here) identified 12 semantic scales that are relevant for describing automotive HVAC sounds (listed in Table 1).

Table 1 – List of the semantic scales. The French scales, in the leftmost columns, were used. The rightmost columns show attempted English translations.

| | | | |
|---|---|---|---|
| Très désagréable | Peu désagréable | Very unpleasant | Not very unpleasant |
| Très sifflant | Peu sifflant | Very whistling | Not very whistling |
| Aigu | Grave | Sharp | Deep |
| Fluctuant | Stable | Fluctuating | Stable |
| Fort | Faible | Loud | Soft |
| Très rond | Peu rond | Very rounded | Not very rounded |
| Très bruité | Peu bruité | Very noisy | Not very noisy |
| Très soufflant | Peu soufflant | Very blowy | Not very blowy |
| Clair | Sourd | Clear | Dull |
| Diffus | Localisé | Diffuse | Localized |
| Lointain | Proche | Far | Near |
| Mauvaise qualité | Bonne qualité | Bad quality | Good quality |

### 2.1 Stimuli

The sound dataset of this experiment must respond to 2 questions: we are interested in the evaluation of automotive HVAC sounds as they are perceived inside the passenger compartment, but also in an evaluation over which the HVAC manufacturers can have an influence. For this reason, the sound dataset is composed of 40 binaural sounds that were recorded in vehicle, which then include the effects of the dashboard and the passenger compartment, and 20 binaural sounds of systems recorded on a test bench in a semi-anechoic room.

Moreover, in order to include a variety as large as possible of HVAC systems, the sound database of the project CESAM (15) was also used. The 60 sounds were selected in order to also include several recordings of each of the following operating modes, at different airflow rates:

- CVAF: cold fresh air emitted at the dash vents (300 kg/h and 500 kg/h)
- CVAR: cold recycled air emitted at the dash vents (300 kg/h and 500 kg/h)
- HDF: warm fresh air emitted at the defrost vent (300 kg/h and 400 kg/h)

- HFF: warm fresh air emitted at the feet vents (300 kg/h and 400 kg/h)

For each sound, a 5-second extract was selected over which linear fade-in and -out of 200 ms were applied. The sounds were also equalized within the frequency band between 80 Hz and 16 kHz in order to compensate for the responses of the artificial heads used for the recordings (Cortex MKII and B&K 4100) and the headphones used in the experiment (Sennheiser HD 650).

### 2.2 Participants

Twenty participants (10 men / 10 women, aged between 23 and 50, car owners) took part in the experiment, for which they were payed. None of them mentioned any major audition problem, and none of them worked in a field related to sound or acoustics.

### 2.3 Apparatus

Sounds were played over headphones in an audio booth in the Laboratoire de Mécanique et d'Acoustique in Marseille. An RME Fireface UC soundcard and Sennheiser HD 650 headphones were used.

A dedicated graphical interface was designed and programmed in Matlab R2011b. This program handled sound playback (with the PsychPortAudio function of the Psychtoolbox library[2]), participants' response input (evaluation of each sound over the 12 semantic scales) and temporary (in case of technical problem) and final response data backup.

### 2.4 Procedure

At the start of the experiment, participants were given written instructions explaining the context of the experiment and the task to perform. After reading these instructions, participants had first to listen to a sequence made of 12 samples extracted from the sound dataset played in a random order for each participant.

After this playback (that could be restarted as many times as desired), the semantic differential experiment started. For each presented sound, participants had to give a rating over each of the 12 7-point semantic scales, before being allowed to evaluate the subsequent sound in the same manner. Participants were given the possibility to listen to the sound and modify its ratings as many times as they wanted before validating. The experiment started with 4 training sounds that were not part of the 60 sounds of the dataset. After rating these 4 sounds on the 12 scales, the experiment continued with the 60 sounds of the dataset. The order of presentation of the 60 sounds was random[3] and different for each participant. Furthermore, the order of presentation of each of the 12 semantic scales and their directions (order of the opposite pair of terms) were random for each participant, but maintained throughout all training and test sounds.

Finally, when the participants had rated all the sounds the experiment was over.

## 3. Analysis

The output data of the experiment is a set of 20 matrices corresponding each to the results of one participant. The size of these matrices is 60 by 12 with the ratings of each of the 60 sounds over each of the 12 semantic scales (only the second presentation of the repeated sound was kept).

Analyses of repeatability and inter-participant consistency – inter-participant correlation matrix and hierarchical analysis with the UPGMA method (19) – revealed occasional inconsistencies but none of the participants had unchangingly outlying responses. No participant was thus removed from the panel for the rest of the analysis.

Considering the fact that the sound dataset included sounds from both vehicle recordings and semi-anechoic recordings, the next step of the analysis consisted in testing whether the recording condition has an effect on the ratings, and if so, over which scales. Twenty sounds among the 60 of the dataset correspond to recordings that were repeated in both conditions, all other thing being equal (same system, operating mode, and airflow rate). A Friedman test was then applied to these 20 sounds for each of the 12 scales. A Bonferroni correction was included in order to compensate for the increase of the Type I error rate due to multiple comparisons. A significant effect of the recording condition was

[2] http://psychtoolbox.org/

[3] However, the first sound presentation was repeated at the end of the experiment in order to check for repeatability. The participants had to rate 61 sounds.

revealed for the scales “very unpleasant/not very unpleasant” ($\chi^2(1,360^4)=12.05$, $p<0.05$), “sharp/deep” ($\chi^2(1,380)=49.54$, $p<0.05$), “loud/soft” ($\chi^2(1,380)=30.57$, $p<0.05$), “far/near” ($\chi^2(1,360^4)=10.52$, $p<0.05$), “very noisy/not very noisy” ($\chi^2(1,380)=25.07$, $p<0.05$), “very blowy/not very blowy” ($\chi^2(1,380)=18.11$, $p<0.05$) and “very rounded/not very rounded” ($\chi^2(1,380)=22.11$, $p<0.05$). Globally, the vehicle recording condition tends to make sounds more unpleasant, louder, deeper, more rounded, nearer, noisier and blowier than the semi-anechoic condition. Only some sounds – CVAR and high-airflow CVAF recordings – are less sensitive to these effects.

In order to identify the links that exist between the semantic scales, a Principal Component Analysis (PCA) was applied to the mean ratings of the 60 sounds over the 12 semantic scales. The first 3 identified components explain more than 95 % of the variance in the data. Figures 1 and 2 respectively show the planes of components 1 and 2, and 1 and 3.

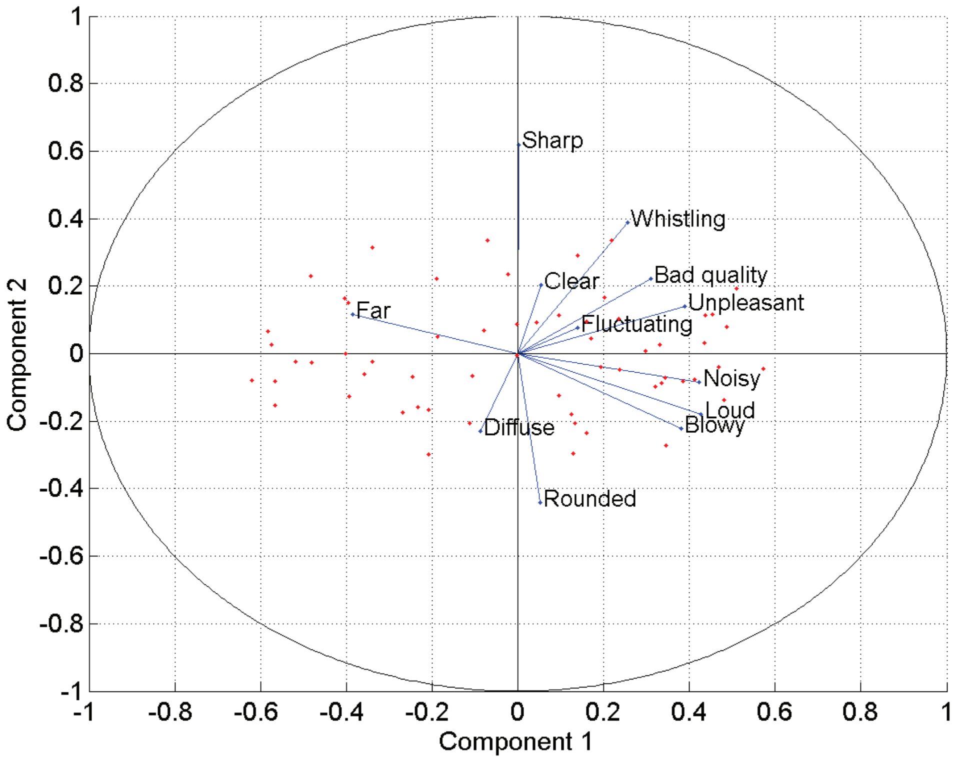


Figure 1 – Plane of components 1 and 2 of the PCA. Red dots represent the sounds; blue lines represent the coefficients of the semantic scales.

The 1st component appears to roughly correspond to unpleasantness ratings (scale “very unpleasant/not very unpleasant”), which are also correlated to the scale “bad quality/ good quality” (r(58 degrees of freedom)=0.95). This component seems also related to loudness (scale “loud/soft”), which is also correlated to the scales “very noisy/not very noisy” (r(58)=0.97) and “very blowy/not very blowy” (r(58)=0.98). Scales “very unpleasant/not very unpleasant” and “loud/soft” are also strongly correlated (r(58)=0.91). Scales “very whistling/not very whistling” (r(58)=0.86), and to a smaller extent “fluctuating/stable” (r(58)=0.80), also seem to have an impact on unpleasantness, and scale “far/near” has an opposite contribution (r(58)=-0.90).

The 2nd component seems linked to the perception of sharpness, since it is aligned with scale “sharp/deep” on the plane of the 2 first components. This scale is also negatively correlated to scale “very rounded/not very rounded” (r(58)=-0.85).

Finally, the 3rd component appears to mainly relate to scale “fluctuating/stable”, which is correlated to scale “very unpleasant/not very unpleasant”, as already mentioned, but also to scale “bad quality/good quality” (r(58)=0.82).

---

[4] Data for this scale had a missing value because of an isolated technical problem during the experiment, which explains why the number of degrees for the error term is lower than for other scales.

Higher-order components can be associated with scales "very whistling/not very whistling" and "diffuse/localized". At last, the scale "clear/dull" seems rather atypical, its higher correlation coefficient, with scale "sharp/deep", being quite moderate (r(58)=0.66).

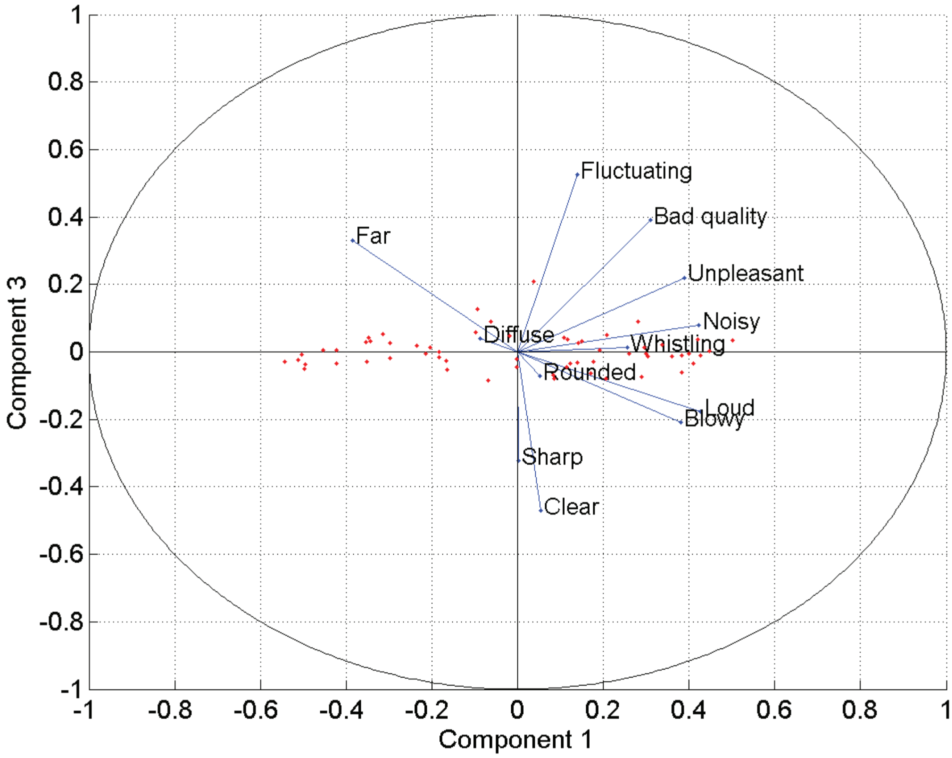


Figure 2 – Plane of components 1 and 3 of the PCA. Red dots represent the sounds; blue lines represent the coefficients of the semantic scales.

The final stage of the analysis consisted in the objectification of the mean ratings on scale "very unpleasant/not very unpleasant" (referred to as unpleasantness for the sake of brevity) through the calculation of acoustic and/or psychoacoustic indicators. For this, the available indicators in the LEA software[5], listed in table 2, were calculated on the 60 sounds of the experiment.

Table 2 – List of calculated acoustic and psychoacoustic indicators

| Indicator type | Details |
|---|---|
| Sound levels | dB SPL; A-, B-, C-, and G-weighting; SEL ; SIL3, SIL4, PSIL |
| Loudness and loudness level indicators | Zwicker's model (20) and Moore's model (21) |
| Other psychoacoustic indicators | Sharpness, Roughness, Fluctuation Strength, Tonality |
| Tonal emergence indicators | ISO1996-2 Annex C (22) and DIN45681 (23) standards |

The adjusted Pearson product-moment correlation coefficient was calculated between the values of each indicator and the unpleasantness scale over the 60 sounds of the dataset. Results show the prominence of loudness in the unpleasantness ratings: the adjusted correlation coefficients of the 2 loudness models are respectively $r_{aj}(58)=0.93$ and $r_{aj}(58)=0.94$, as well as those of the A-weighted

[5] http://www.genesis-acoustics.com

sound level ($r_{aj}(58)=0.94$) and the Speech Interference Levels (SIL3, SIL4 et PSIL, with $r_{aj}(58)=0.94$, 0.94 and 0.91 respectively). The other weighted or unweighted sound levels are less efficient and the other psychoacoustic indicators prove to be individually insufficient for explaining the unpleasantness ratings.

In order to improve the prediction of the unpleasantness, stepwise regression (24) was used to find a more comprehensive sound quality model. This procedure iteratively searches for the indicators that significantly improve the model prediction. It resulted in a model combining the loudness level according to Moore's model and Tonal Audibility of standard ISO1996-2: $r_{aj}(57)=0.96$, $F(57,3)=337.49$, $p=0.00$ (see Figure 3). However, the contribution of loudness level is more than 5 times as high as that of Tonal Audibility. Other models of sound quality for HVAC systems were also tested: Zwicker and Fastl's Psychoacoustic Annoyance (25), the pleasantness model of project CESAM (15), based on SIL4 and Aures' tonality, and Menager and Rochepeau's preference model (14), which was defined for a former generation of automotive HVAC systems and is based on loudness level (Zwicker's model) and PSIL (this 2 indicators, however, are highly correlated with one another over the 60 sounds of the dataset, $r_{aj}(58)=0.98$). These models do not improve significantly the prediction of the model: their correlation coefficients are respectively $r_{aj}(58)=0.90$, -0.94 et -0.93.

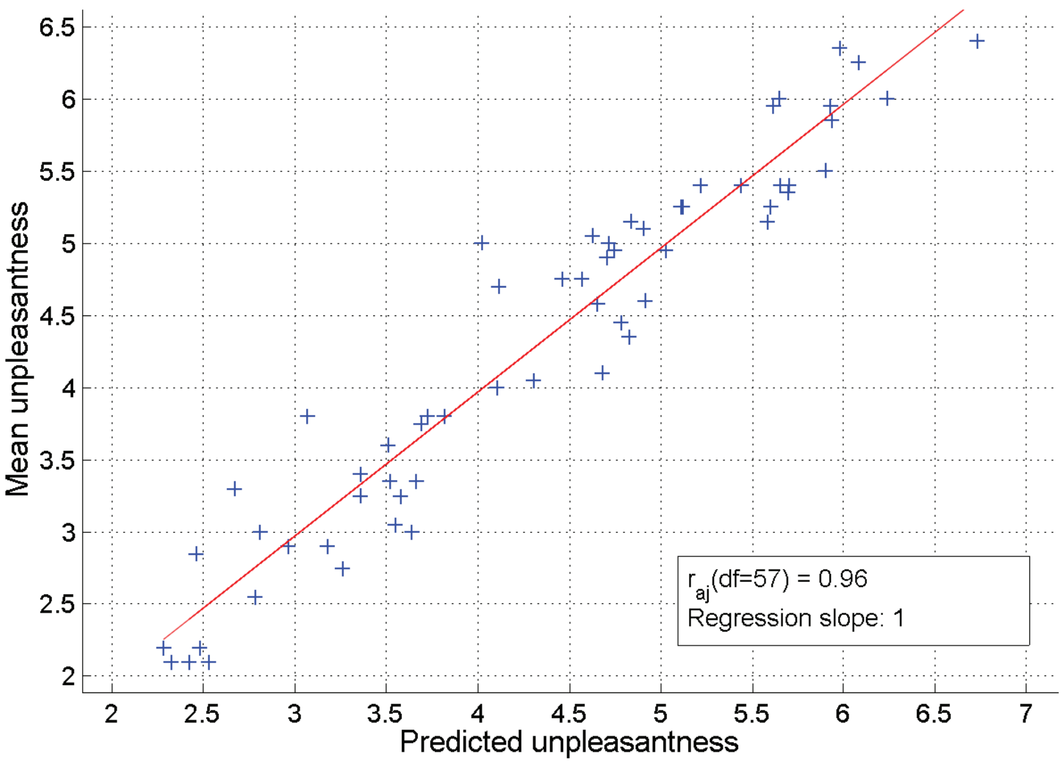


Figure 3 – Scatter plot of the mean values of unpleasantness and the predicted unpleasantness values on the basis of Moore's loudness level and ISO1996-2 Tonal Audibility

## 4. Discussion

The unpleasantness model identified in this study must be looked at in regards to models identified in former studies on HVAC sounds, whether in the automotive field or not.

First this model is quite similar to that of project CESAM (15) that includes a sound level indicator (SIL4) and an indicator partly based on a tonal emergence calculation (Aures' tonality). The model from (14) is somewhat different since it is based on 2 sound level indicators (loudness level and PSIL), that happen to be correlated with one another for the studied sound dataset.

The stepwise regression in this study initially included the G-weighted sound level. At first glance, this was in line with a former study on residential HVAC systems (10), which revealed a contribution of the sound energy in low frequencies described by the specific loudness in the 2 first Bark bands.

However, the G-weighted sound level was not maintained in the model for different reasons. First the correlation coefficient in the stepwise regression in this study appeared to be negative, which would mean that a higher G-weighted sound level would decrease unpleasantness. This seems inadequate, and noticeably in contradiction with the model in (10). Moreover, the G weighting is centered on 20 Hz. It is rather unlikely that this part of the spectrum had an influence during the listening test, because of the headphones that were used and since no subwoofer was used. Its presence in the model is thus most probably incidental.

It may also seem surprising not to find in the regression analysis other psychoacoustic indicators often associated to HVAC sounds, such as sharpness or fluctuation strength. The principal component analysis revealed that the 2nd and 3rd components were indeed associated to scales “sharp/deep” and “fluctuating/stable”, respectively. The previously mentioned study (10) also identified sharpness as an auditory attribute of the timbre of HVAC sounds by means of dissimilarity measure experiment, but its contribution to preference was uncertain (noticeably the stepwise regression performed in that study did not include a sharpness indicator). In another study (9) acoustic comfort was directly linked to sharpness (a higher value had a detrimental effect), but the conducted experiments only compared sounds corresponding to the same recordings whose tonal components and a specific frequency band were either amplified or reduced. In that case correlation between sharpness and acoustic comfort is implied by the experimental design. In the end, the influence of sharpness on sound quality is not trivial even though sharpness is a sound parameter that is clearly perceived. It is also possible that the link between sharpness and sound quality be neither monotone nor linear.

As for fluctuation strength, its influence is probably too occasional in the studied sound dataset (mainly one single sound among 60, which corresponded to a vehicle that was not recent and whose HVAC system appeared to be deteriorated) to be included in the objectification model. While unpleasantness seems to be much higher when strong fluctuations appear, both the rarity of this phenomenon in the sound dataset and the fact that it was caused by deterioration indicates that taking into account fluctuation strength is not required when considering a model of sound quality of out-of-factory HVAC systems.

## 5. Conclusion

The goal of this study was to study the perception of automotive HVAC systems and more specifically the unpleasantness that they cause. A sound dataset was defined in order to include a large variety of HVAC systems, from either combustion, electric or hybrid vehicles. Relevant auditory attributes for describing HVAC sounds were identified through a psychoacoustic experiment of semantic differentials. In this experiment 60 HVAC sounds were rated over 12 semantic scales that were defined during a preliminary verbalization experiment. A multidimensional analysis (PCA) revealed a perceptual space mainly composed of 3 dimensions respectively associated to unpleasantness, sharpness and fluctuation. The objectification of the unpleasantness scale resulted in a prediction model based on loudness level and an indicator of tonal emergence (tonal audibility of standard ISO1996-2). This model appears to be slightly more efficient than former models of automotive HVAC sound quality.

This study, along with former studies on the same topic (14, 15), is based on a psychoacoustic approach to sound quality. Nevertheless, in the context of trains, the studies of Khan (12) and Mzali (13) have already shown that a cognitive approach of sound quality can reveal a beneficial impact of HVAC sounds. Such an approach could also be used for automotive HVAC sounds, and would involve the evaluation of the sound ambiance inside a car passenger’s compartment while driving. Thus the sound of the HVAC system would be perceived along with other sound sources (engine noise, rolling noise, buffeting noise, etc.), and its study could result in a new definition of the sound quality of automotive HVAC systems embedded in a more general concept of acoustic comfort inside cars. Such results could prove essential for the design of optimal HVAC systems adapted to new automotive technologies.

## ACKNOWLEDGEMENTS

CEVAS is a FUI AAP 14 project within the framework of the I-TRANS and PEGASE clusters. The involved partners are VALEO, CETIM, University of Technology of Compiègne, ESI GROUP and GENESIS. The GENESIS’ work described here was funded by FEDER, BPI France and CPA (Communauté du Pays d’Aix-en-Provence).

## REFERENCES

1. Bennouna S, Ouedraogo B, Moreau S, Ville JM, Chériaux O. Aeroacoustic Measurement of Automotive HVAC In-Duct Elements. Proc FAN 2015, Lyon, France 2015.
2. Legros M, Ville JM, Moreau S, Carniel X. Acoustic Synthesis of an Automotive HVAC. Proc FAN 2015, Lyon, France 2015.
3. Minard A, Vidal A, Lambourg C, Boussard P. Sound Synthesis of Fan Noise and Modeling of its Perception in Car Passenger Compartment. Proc FAN 2015, Lyon, France 2015.
4. Ko NWM, Ho WF, Un WK. Responses to air-conditioning system noise. J Sound Vib. 1978;57(4):595-602.
5. Tang SK. Peformance of noise indices in air-conditioned landscaped office buildings. J Acoust Soc Am. 1997;102(3):1657-1663.
6. Tang SK, Wong MY. On noise indices for domestic air-conditioner noise. J Sound Vib. 2004;274(1-2):1-12.
7. Ayr U, Cirillo E, Martellotta F. An experimental study on noise indices in air conditioned offices. Appl Acoust. 2001;62(6):633-643.
8. Susini P, McAdams S, Winsberg S, Perry I, Vieillard S, Rodet X. Characterizing the sound quality of air-conditioning noise. Appl Acoust. 2004;65(8):763-790.
9. Jeon JY, You J, Chang HY. Varying the spectral envelope of air-conditioning sounds to enhance indoor acoustic comfort. Build Environ. 2011;46:739-746.
10. Minard A. Perception et confort acoustiques des systèmes de traitements d'air. PhD thesis, Université de La Rochelle, 2013.
11. Khan MS, Hogstrom C. Determination of sound quality of HVAC systems on trains using multivariate analysis. Noise Control Eng. 2001;49(6):276-283.
12. Kahn MS. Effects of masking sound on train passenger aboard activities and on other interior annoying noises. Acta Acust united Ac. 2003;89(4):711-717.
13. Mzali M. Perception de l'ambiance sonore et évaluation du confort acoustique dans les trains. PhD thesis, University Paris VI, 2002.
14. Menager M, Rochepeau D. Etude de la Qualité Sonore des appareils de chauffage et de climatisation automobile. Technical report, INSA pour Valeo, 2004.
15. Marteau J, Besombes M. Jury d'écoute sur bruits de HVAC. Rapport de la tâche 2.7.4 du projet REVA-CESAM "Extraction de critères psychoacoustiques", 2009.
16. Aures W. A model for calculating the sensory euphony of various sounds. Acustica. 1985;59:130-141.
17. Solomon LN. Semantic approach to the perception of complex sounds. J Acoust Soc Am. 1958;30(5):421-425.
18. Nosulenko V, Parizet E, Samoylenko E. La méthode d'analyse des verbalisations libres : une application à la caractérisation des bruits de véhicules. Informations sur les Sciences Sociales. 1998;37(4):563-611.
19. Legendre P, Legendre L. Numerical ecology. 3e ed., Elsevier, 2012.
20. ISO532:1975 (R2012), Acoustics -- Method for calculating loudness level, 1975.
21. ANSI S3.4-2007, American National Standard Procedure for the Computation of Loudness of Steady Sounds, 2007.
22. ISO1996-2:2007, Acoustics - Description, measurement and assessment of environmental noise - Part 2: Determination of environmental noise levels, 1987.
23. DIN45681:2005-03, Acoustics – Detection of tonal components of noise and determination of tone adjustment for the assessment of noise immisions, 2005.
24. Howell DC. Statistical method for psychology. Wadsworth Publishing, 2007.
25. Zwicker E, Fastl H. Psychoacoustics: Facts and Models. Springer, 1990.