SOUND SYNTHESIS OF FAN NOISE AND MODELLING OF ITS PERCEPTION IN CAR PASSENGER COMPARTMENT

Antoine MINARD, Adrien VIDAL,
Christophe LAMBOURG, Patrick BOUSSARD

*GENESIS Acoustics, Domaine du petit Arbois, BP 69,
13045 Aix-En-Provence Cedex 4, France*

SUMMARY

In the context of the CEVAS project which addresses the problem of modelling the sound produced by either air conditioning or electric battery cooling systems in cars, the present work deals with the synthesis of such sounds as they are perceived by the passengers. A sound synthesis algorithm was developed in order to generate realistic sounds at the passengers' ears positions and validated through a perceptual experiment. A methodology for acoustically characterising inner car compartments was developed and applied to free-field recordings of HVAC through auralisation in order to reproduce realistic HVAC 3D sound within the passenger compartment.

INTRODUCTION

The CEVAS project addresses the problem of modelling the sound produced by either air conditioning (HVAC) or electric battery cooling (BTM) systems from the definition of the sound sources to their perception by users. The main aim of the project is to build a reliable and efficient tool for designing new systems according to high quality standards in terms of acoustic comfort.

The other partners in the CEVAS project are interested in the aero-acoustic characterisation and modelling of car HVAC and BTM systems that will provide spectral description of the emitted sound, to different levels of precision. For the sake of completeness, GENESIS' work in the project addresses the questions of how car HVAC and BTM systems sound, how their sounds are modified within the passenger compartment, and how they are eventually perceived by passengers. Thus, the part of GENESIS is to generate the sound signals that will be perceived at the passengers' ears from the spectral data, and to study how people react to these sounds in terms of acoustic comfort. The attainment of these goals entails (i) the design of a tool for synthesising an audio signal corresponding to acoustic spectra defined on the dashboard, (ii) the development of a methodology for reproducing the sound propagation inside the passenger compartment, and (iii) the study of the resulting sound perception by users. This paper focuses on the two first points, for HVAC systems only. It was hypothesised that this work could be later transposed to BTM systems.

In the end, the aim of the work presented here is to be able to reliably reproduce the sound produced by a HVAC system through headphones as it would be perceived by a user in a real passenger compartment. Such a tool will allow manufacturers to listen to the sound produced by HVAC systems only from virtual designs in a realistic manner.

SYNTHESIS OF HVAC SOUNDS AND PERCEPTUAL VALIDATION

The development of a tool for synthesising car HVAC sounds first needs the gathering of a sound recording database of existing car HVAC systems. After the creation of this database, the recorded sounds were analysed in order to generate spectral data with different precisions. These data served as an input for an additive synthesis tool in order to reproduce the initial sounds according to the spectra definitions. Finally the perceptual fidelity of the synthesised sounds to the original was compared through a listening test.

HVAC sound recording

The recordings come from 2 databases. The first one corresponds to sounds specifically recorded in the framework of the CEVAS project. The second one corresponds to recordings from a former project on car HVAC systems, the REVA-CESAM project.

- CEVAS database: sounds from a single HVAC system dismounted off the dashboard were recorded at a sampling frequency of 65536 Hz in VALEO semi-anechoic room (Verrière, France) with a Brüel & Kjær 4190 pressure microphone and a Brüel & Kjær 4100 dummy head, both associated to a Brüel & Kjær 2669 amplifier. Four common operating modes were recorded at 2 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 vent (300 kg/h and 400 kg/h)
- CESAM database: the sounds are described in the technical report of the REVA-CESAM project [1]. They were recorded at a sampling frequency of 48 kHz with a Brüel & Kjær 4100 dummy head. They correspond to different HVAC systems from several brands operating in either CVAF or CVAR mode and at an airflow rate of either 300 or 500 kg/h. Moreover, recordings were subsequently high-pass filtered with a cutoff frequency of 150 Hz in order to remove low-frequency noise created by air turbulence around the dummy head pinnae.

All sounds were eventually resampled at a frequency of 44.1 kHz and their durations were set to 5 s.

Recording analysis and additive synthesis

The used synthesis principle is to consider the HVAC sound as a broadband noise with possible additional tonal emergences (*additive synthesis* [2]). Here the sounds were considered as stationary, which appeared as a natural hypothesis when listening to the sounds of both databases. The tested analysis-synthesis models differ in the precision of the spectral data used as an input of the synthesis, and whether or not additional tones were included.

The analysis of the signals consisted in extracting the input data for the synthesis models:

1. The Power Spectral Density (PSD) of the broadband noise was estimated by Welch's method (Hanning window with 50% overlap), with varying DFT[1] numbers of points (256, 512, 1024, 2048 and 4096).
2. Octave-band (10 bands from 31 Hz to 16 kHz) and 1/3-octave-band (29 bands from 25 Hz to 16 kHz) levels were calculated through time-domain filtering.
3. When a tone was emerging, its frequency was estimated by searching for a local maximum of the PSD (with 16384 DFT points) and its level was evaluated by summing the energies of the located peak and the two closest frequency bins.

Note that when a tone was emerging the broadband noise PSD and band levels were estimated without the tone energy contribution.

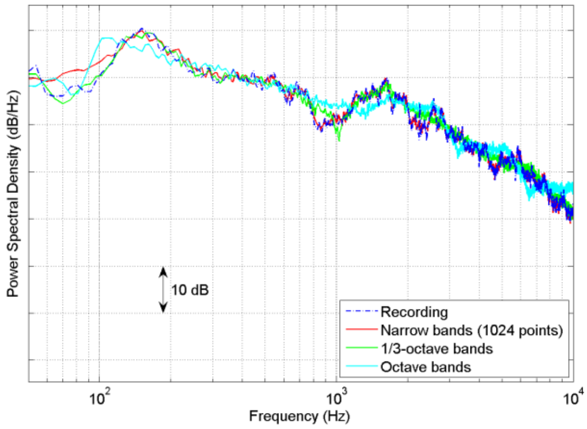

Figure 1: PSD (Welch method with 8192 DFT points) of a CEVAS recording and the results of the corresponding syntheses with different spectral descriptions

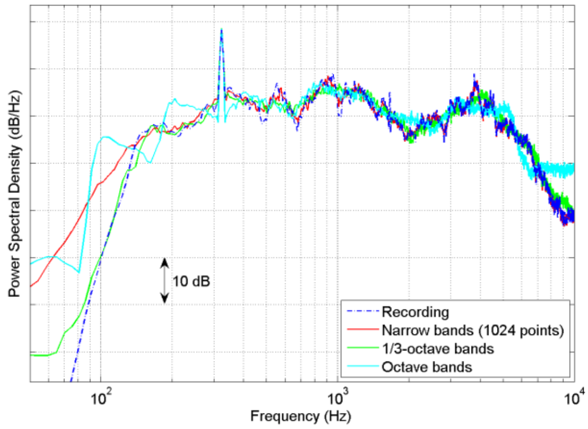

Figure 2: PSD (Welch method with 8192 DFT points) of a CESAM recording presenting a tonal emergence and the results of the corresponding syntheses with different spectral descriptions and additional sinusoidal tones

The sound synthesis consisted in transforming these spectral data into a time-domain signal and was performed by GENESIS' GenePASS software. The broadband noise was synthesised by the Overlap Add method [3]. Two examples of resulting syntheses are shown on Figure 1 and Figure 2, respectively for a CEVAS recording with only broadband noise and a CESAM recording with an additional emerging tone. Three resolutions of spectral description were considered, narrow bands (PSD with 1024 DFT[2] points), 1/3-octave bands and octave bands. The displayed spectra were calculated with 8192 DFT points, which is a higher resolution than all spectral analyses described above, and thus enabled a thorough comparison. Note that although synthesis from only narrow-band description is theoretically able to reproduce emerging tones it requires a highly resolved spectral description, which 1024 DFT points may not allow. This is the reason why we also considered adding sinusoidal signals to broadband noises synthesised form narrow-band analysis.

Perceptual validation

The selection and validation of the suitable synthesis model aims at defining the precision of the input data necessary for a reliable reproduction of recordings. As a consequence, we need to

[1] Discrete Fourier Transform.

[2] When informally listening to the sounds, this number of DFT points seemed to enable almost indistinguishable syntheses and recordings (at least when no tone was emerging) and was thus selected out of other narrow-band descriptions for further analyses.

perceptually compare the results of different syntheses of the same recordings. The idea is not to identify which model enables an indistinguishable reproduction of the recordings, but rather to verify that the distribution of the timbres of the synthesised sounds correspond to that of the recordings timbres. In other words, the perceptual dissimilarities between syntheses and the corresponding recordings need to be significantly smaller than those between recordings.

This specific experimental positioning ruled out procedures such as ABX test, because we already expected that differences could be easily perceived between syntheses and recordings. Since we were also interested in comparing recording-synthesis with between-recordings perceptual dissimilarities, MUSHRA-type procedure were not suitable either. Measuring perceptual dissimilarities rather points towards timbre-related procedures, such as similarity measurement, or even free-sorting task. Finally a so-called "pick-any" procedure, first introduced by Rao & Katz [4], was implemented.

In this procedure, participants have to select k sounds among p presented test sounds that seem the most similar to a reference sound. Each sound as reference must be confronted to all others, which means that, for a total of N sounds, $N(N-1)/p$ trials (as compared to $N(N-1)/2$ trials for a standard paired-comparison procedure) are presented to participants. For each trial, the recorded data take the form of a binary value for each of the p pairs of reference/test sounds indicating whether or not the test sound was selected. These binary values can then be placed in a square matrix where each cell corresponds to a pair of sounds. Through averaging over the panel of participants, the resulting matrix can be regarded as a (dis)similarity matrix. The advantage of this procedure is that it allows us to build a perceptive space of a sound dataset (through multidimensional analysis) with fewer trials than usual methods when $p > 2$. In a study on the sound reproduction quality of loudspeakers [5], Michaud found out that the results obtained with this procedure (with $p = 3$ and $k = 1$) were consistent with those obtained with similarity measurement and free-sorting task procedures. In the present case we set $p = 4$ test sounds and $k = 2$ sounds to choose, in order to avoid a constant bias in the dissimilarities. As a consequence the sound dataset size must be a number equal to $4n + 1$ (with any integer n).

- *Stimuli*: Initially our intention was to associate into a single database the CEVAS and CESAM recordings. However the use of a high-pass filter at 150 Hz in the latter might have yielded highly categorical results that would prevent an efficient comparison of the synthesis/recording dissimilarities with the between-recordings dissimilarities. Hence we considered 2 separate sound datasets for the experiment. The CEVAS sound dataset included 13 sounds:
  - 4 recordings (labelled '**rec1**' to '**rec4'**), corresponding to 4 operating modes of the same HVAC system,
  - for each recording a synthesis from the 1024-point narrow-band spectrum ('**syn<i>-nb**'),
  - for each recording a synthesis from the 1/3-octave-band levels ('**syn<i>-3b**'),
  - a duplicate ('**rec2d**') of one of the recordings because of the required number of sounds for the procedure.

The CESAM sound dataset included 17 sounds:

- 4 recordings (labelled '**rec5**' to '**rec8'**), corresponding to 4 different HVAC systems (and varying operating modes),
- for 2 of these recordings ('**rec5**' and '**rec6**'), the 2 same synthesis types as in the CEVAS dataset,
- for the 2 others, the 2 syntheses are also reproduced each with an additional sine modelling the tonal emergence in the corresponding recordings ('**syn<i>-3b+t**' and '**syn<i>-nb+t**'),

- a 16384-point narrow-band synthesis (‘**syn8-nb16**’ enabling the reproduction of the tone) because of the required number of sounds for the procedure.

Finally, the sounds were loudness-equalized (10.1 sones according to Zwicker’s model [6]) so that the judgements would not be impacted by the perceived sound level. The sounds were 5 s long with linear fade in and out of 0.2 s.

- *Apparatus*: A dedicated MATLAB R2011b graphical user interface (Figure 3) including trial series control, sound playback and data recording was developed. Sounds were played over Sennheiser HD650 headphones through an RME Babyface audio card.

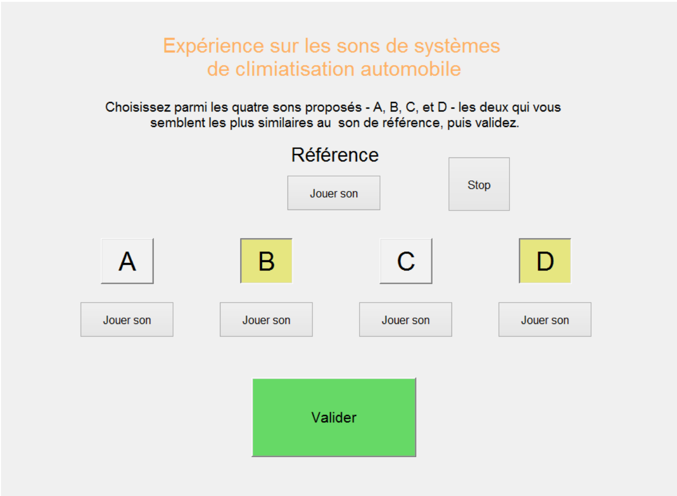

Figure 3: Graphical User Interface for the experiment

- *Participants*: Twelve participants (10 men, 2 women, aged between 22 and 47 years old) volunteered as listeners. All reported having normal hearing.

- *Procedure*: The same procedure was reproduced for both sound datasets. At the beginning of the experiment, participants were given written instructions. All 5-sound groups (reference and test sounds) for each participant were created beforehand. The stimuli presentation order was randomly permuted although it was verified that the successive presentations of each sound was sufficiently spread over the whole test. The presentation order also included one training trial and one repeated trial for repeatability analysis. The experiment included 41 and 70 trials respectively for the CEVAS and the CESAM datasets.

- *Results*: The mean durations of the experiment were 25 and 26 minutes respectively for the CEVAS and the CESAM datasets. The experiment was easier for the CESAM dataset because of the diversified syntheses (additional tone) which explains the comparable durations, as opposed to different numbers of trials. For each datasets, the experimental data consist of individual binary matrices (where each cell (i, j) is 1 if the sound j was selected when compared to reference sound i, and 0 otherwise) that were averaged to form a cooccurrence matrix. The cooccurrence matrix represents how many participants have associated each pair of sounds by selecting one when the other was the reference sound. This can also be interpreted as a proximity matrix [7].

    - CEVAS sound dataset

Figure 4 shows the cooccurrence matrix obtained for the CEVAS sound dataset. On this figure, the brighter a cell is the more similar the corresponding pair of sounds is. The brightest parts of the matrix (close to the diagonal) correspond to pairs of recording/synthesis or synthesis/synthesis of the same recording. This is consistent with our starting hypothesis which is that recordings and respective syntheses are more similar than pairs of recordings.

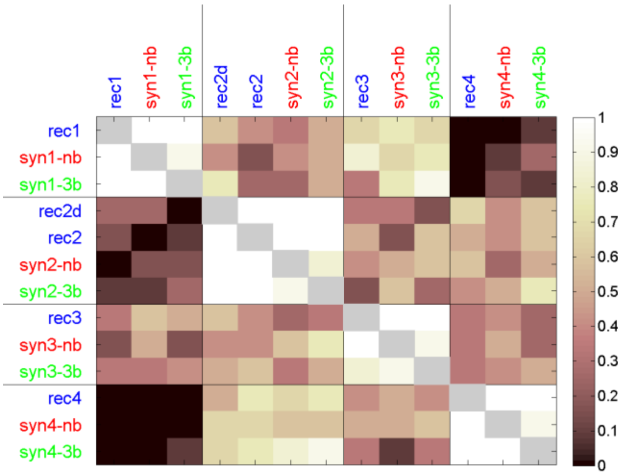

Figure 4: Similarity matrix for the CEVAS sound dataset. Recordings are indicated in blue, narrow-band syntheses in red and 1/3-octave-band syntheses in green

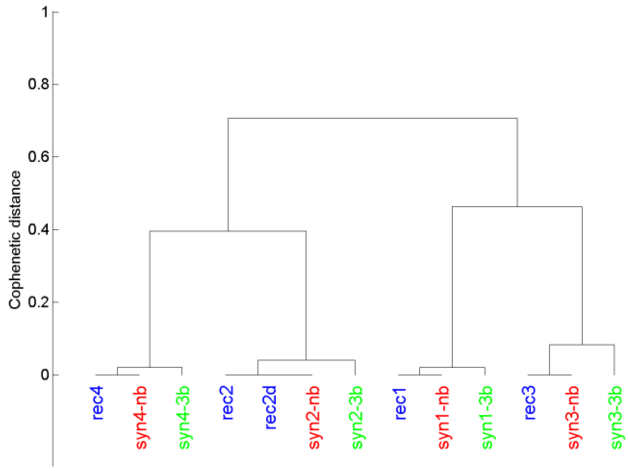

Figure 5: Dendrogram representation of the CEVAS sound dataset. Recordings are indicated in blue, narrow-band syntheses in red and 1/3-octave-band syntheses in green

In order to confirm this trend we derived a hierarchical tree representation (*dendrogram*) from these data using an unweighted arithmetic average clustering (UPGMA) analysis algorithm (see [8] for computational details). In such a representation, the perceptive distance between two sounds is modelled as the height of the node that links them (*cophenetic distance*). The obtained dendrogram is shown on Figure 5. It can be clearly seen that the sounds are gathered into 4 groups, each with one of the recordings and the corresponding syntheses. Moreover in all groups there are slight distinctions between the 1/3-octave-band syntheses and both the recordings and the narrow-band syntheses, but probably too small to be conclusive.

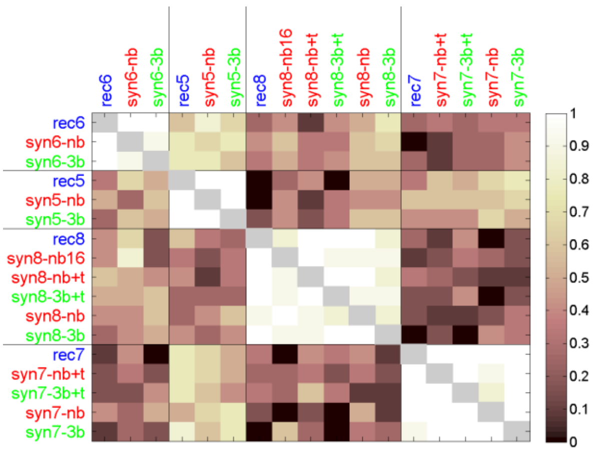

Figure 6: Similarity matrix for the CESAM sound dataset. Recordings are indicated in blue, narrow-band syntheses in red and 1/3-octave-band syntheses in green

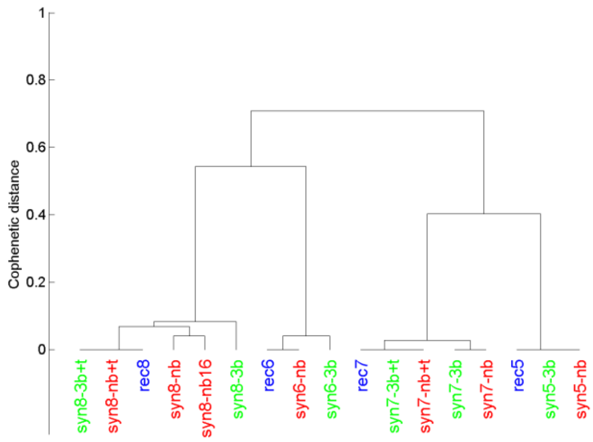

Figure 7: Dendrogram representation of the CESAM sound dataset. Recordings are indicated in blue, narrow-band syntheses in red and 1/3-octave-band syntheses in green

- CESAM sound dataset

The same analysis was applied to the experimental data for the CESAM sound dataset. Figure 6 shows the resulting similarity matrix. The same trend as for the CEVAS sound dataset can be observed: the recordings and the respective syntheses are more similar than pairs of recordings. Figure 7 shows the corresponding dendrogram and consistent results (grouping of each recording with its corresponding syntheses). However, whereas no obvious synthesis efficiency improvement can be observed between 1/3-octave-band and narrow-band syntheses, the consideration of the additional tone (‘**syn<i>-nb+t**’ and ‘**syn<i>-3b+t**’) seems to induce a synthesis slightly more similar to the recording than the mere broadband noise synthesis.

Discussion

Narrow-band synthesis appears to be the most efficient model. However, for sounds with an emerging tone, separate sine synthesis seems also important, although the conclusiveness of this trend suffers a rather small number of involved examples. Besides, 1/3-octave-band synthesis results in perceivable differences with recordings, but seems able to reproduce the main timbre characteristics of the HVAC recordings nonetheless.

In the end this part of the study provided an efficient tool for synthesising HVAC sounds, allowing a rather accurate sound rendering from only broadband spectral description and frequency and magnitude of emerging tones. Hence this tool will enable reliable audio signal production derived from data obtained by the acoustic modelling of HVAC sound sources conducted by the partners of the project.

SPATIAL SOUND RENDERING THROUGH AURALISATION

Auralisation designates the reproduction of a spatialized virtual sound environment. In the scope the CEVAS project, we need to recreate realistic listening conditions of HVAC systems within car passenger compartment. Here the input is the sound pressure signal emitted towards an arbitrary reference position (1 m in front of the vent) by equivalent sources at the dash vent positions in free-field condition. In the present context the methodology consists in:

- Estimating the transfer functions between the sources and the passengers’ ears.
- Convolving the associated impulse responses with the anechoic signal of the source.
- Summing the contributions of the different sources.
- Playing the resulting signals corresponding to a particular passenger position over headphones.

There are two ways of estimating the transfer functions:

- through numerical acoustic calculation methods (geometrical acoustics modelling and/or Finite Element Method for low frequencies) from a CAD model of a passenger compartment,
- through measurements with a calibrated electroacoustic source.

In this study, the transfer functions were estimated through measurements in a real car passenger compartment. These measurements thus consist in playing a broadband noise through an electroacoustic source placed over the dash vents and recording the responses for both ears of a dummy head at the passengers’ positions (here only the driver’s position is considered).

Design of a dedicated sound source

This part of the project entails the design of an electroacoustic source suited for the subsequent measurements. The specifications are:

- quite thin and compact for the loudspeaker to be close to the vents while limiting diffraction,
- enabling 50-Hz-to-20-kHz transfer function measurements,
- directivity similar to that of the real sources – given the sources nature (small pipe opening compared the wavelength) it is assumed that vents are mainly omnidirectional sources, which also tends to be the case of small enclosed loudspeakers,
- calibrated.

To meet these specifications, we designed an enclosed electrodynamic transducer of suitable shape. The loudspeaker is a 1"-diameter Aurasound NSW1-205-8A with a cutoff frequency of 220 Hz. The enclosure has a volume of 40 cm^3 similar to the equivalent air compliance volume (V_{AS}) of the loudspeaker. Thus the enclosed loudspeaker has a cutoff frequency of 310 Hz. Moreover, in order to limit diffraction, the enclosure was shaped as a spherical cap. It was made through ABS 3D prototyping and the result can be seen on Figure 8.

Figure 8: Photograph of the designed loudspeaker

Characterisation of the electroacoustic source

Measurements were conducted in the semi-anechoic room of the Laboratoire de Mécanique et d'Acoustique (LMA) in Marseille, France. The room has one reflective side wall at a distance of 8 m from the opposite wall wedges.

- *Estimation of the nominal level:*

The nominal level is defined as the highest level a loudspeaker can withstand without creating too much distortion. The distortion is measured by the THD (Total Harmonic Distortion) at different frequencies and input levels, which is the ratio of harmonics and total energies:

$$THD = 100 * \frac{\sqrt{\sum_{k=2}^{N} P_{xx}(k)}}{\sqrt{\sum_{k=1}^{N} P_{xx}(k)}} \quad (1)$$

with P_{xx} is the recorded sine autospectrum and k is the harmonic index. N was set to 40.

Measurements were performed in the centre of the semi-anechoic room with an MK231E microphone and an MV210 preamplifier at a distance of 25 cm from the loudspeaker. The excitation signal of the loudspeaker was amplified by a Trends TA10.1 amplifier. Figure 9 shows the results. A level of -32 dBFS was selected as a trade-off between low THD value and high operating level.

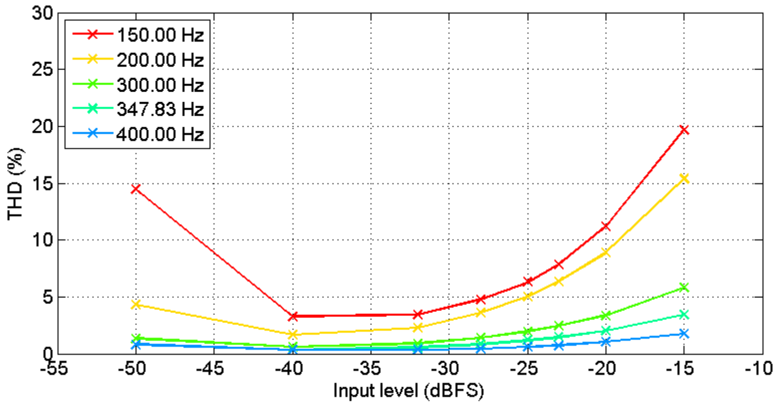

Figure 9: Electroacoustic source THD as a function of input level for different frequencies

- *Directivity and frequency response measurements:*

The loudspeaker was positioned around 7 m away from the reflective wall while facing it. The same microphone as before was placed at a distance of 1 m from the loudspeaker. Measurements with the H_1 estimator were repeated for angles of radiation 0°, 15°, 30°, 45° and 60°. The resulting frequency responses are displayed on Figure 10. The frequency responses are similar between 150 Hz and 6 kHz for all radiation angles. This demonstrates the omnidirectional characteristic of the enclosed loudspeaker within this bandwidth. Below 150 Hz, the deviations may come from the measurements uncertainty (poor signal-to-noise ratio).

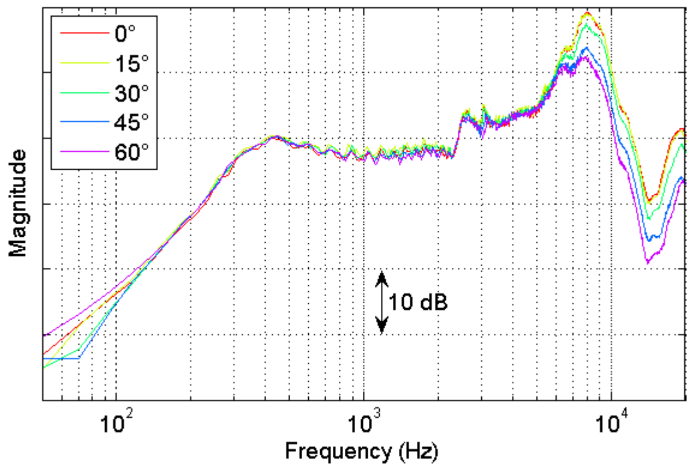

Figure 10: Frequency response of the electroacoustic source for different angles of radiation

Transfer function measurement of car passenger compartment

Now that the electroacoustic source is fully characterised, it can be used to measure transfer functions of the car passenger compartment.

- *Impulse response measurement:*

The measurements were performed on a C-segment station wagon with four air vents on the dashboard (2 in the centre, 2 on the sides) in a quiet sound environment. Measurements were performed with a Cortex Mk2 dummy head and the same measurement hardware as for the loudspeaker characterisation (including the microphones and preamplifiers in the dummy head). Also the same estimator and parameters were used. Figure 11 shows the resulting frequency responses – left (blue) and right (red) ears – for the left side vent, as compared to the frequency response of the electroacoustic source (black).

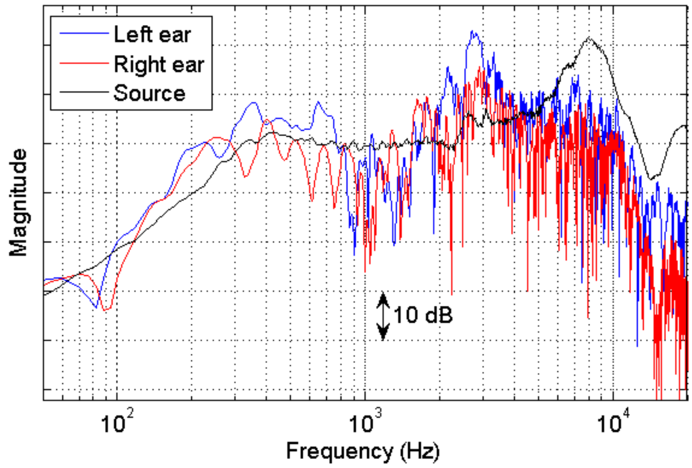

Figure 11: Frequency responses of the passenger compartment

- *Response equalisation:*

The ears frequency responses (blue and red curves) shown on Figure 11 do not take into account the non-uniformity of the electroacoustic source response (black curve). Therefore they need to be adjusted before being applied to anechoic HVAC recordings. Hence an inverse filter was estimated from the electroacoustic source frequency response measured at 0° radiation angle, by using Tikhonov regularization method [9] which guarantees the stability of the filter (FIR filter).

The measured impulse responses of the passenger compartment were then equalised by convolution with the obtained filter. This neutralised the bias introduced by the electroacoustic source and the other elements of the measuring chain. It thus provides the actual responses of the passenger compartment (Figure 12).

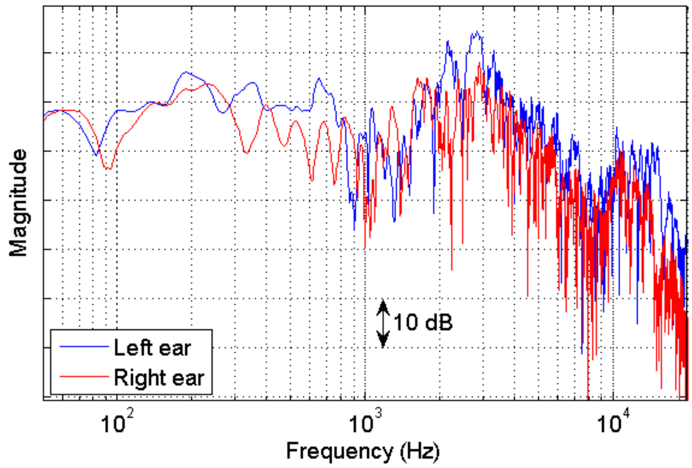

Figure 12: Equalised frequency responses of the passenger compartment

Auralisation of equivalent sound sources

In the final step, the measured impulse responses of the passenger compartment will be convolved with the "free-field" signals emitted by the equivalent sources at the dash vents. This will reproduce the acoustic propagation of the equivalent sources towards the passengers' ears. For now the vents contribution in free-field has yet to be characterised.

In the scope of the CEVAS project, future work will also include the extension of this methodology to the passenger compartments of other existing vehicles of different segments. The results will be compared with binaural recordings of the sounds of embedded HVAC systems performed inside the corresponding passenger compartments.

CONCLUSIONS

In the framework of the CEVAS project, this study allowed us to design and test a reliable tool for synthesising audio signals corresponding to the acoustic spectra of car HVAC systems. In order to reproduce realistic HVAC sound timbres with data usually produced by acoustic modelling from the partners of the project, several spectral resolutions were compared in terms of perceptual dissimilarity to the original sound in a listening test. This revealed that almost indistinguishable reproduction of the broadband noise of the recordings is obtained with synthesis from 1024-point PSD. Synthesis from 1/3-octave-band levels creates some perceivable differences with the original, but the main timbre characteristics are reproduced. However, when a tone is emerging, it needs to be separately synthesised since both 1024-point-PSD and 1/3-octave-band spectral resolutions are unsufficient.

Acoustic propagation inside the car passenger compartment was reproduced through convolution with measured acoustic transfer functions. The transfer functions of a car passenger compartment were characterised by measurements with a specifically designed electroacoustic source. The source was fully characterised so as to counterbalance the bias introduced by its electroacoustic response. The transfer functions were finally applied to anechoic recordings for spatialized sound rendering of HVAC systems. The perceptual efficiency of this procedure yet remains to be studied further by extension to other car passenger compartments.

The last part of the GENESIS' work in the CEVAS project – not addressed in this paper – deals with the psychoacoustic study of the sound quality of car HVAC systems and is currently being conducted. Eventually objectivation of the sound quality of car HVAC systems will be addressed. It consists in predicting listeners' judgments from audio features calculated from the sound signals. Hence it will provide a reliable metric for evaluating the expected sound quality of current or future HVAC systems.

ACKNOWLEDGEMENT

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).

BIBLIOGRAPHY

[1] J. Marteau, M. Besombes – *Jury d'écoute sur bruits de HVAC. Rapport de la tâche 2.7.4 du projet REVA-CESAM "Extraction de critères psychoacoustiques".* Technical report, **2009.**

[2] X. Serra – *A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition.* PhD thesis, Stanford University **1989.**

[3] J. B. Allen, L. R. Rabiner – *A Unified Approach to Short-Time Fourier Analysis and Synthesis.* Proceedings of IEEE , **65**(11), pp. 1558-1564, **1977**.

[4] V. R. Rao, R. Katz – *Alternative multidimensional scaling methods for large stimulus sets.* Journal of Marketing Research, **8**(4), pp. 488-494, **1971**.

[5] P.-Y. Michaud – *Distorsions des systèmes de reproduction musicale : Protocole de caractérisation perceptive.* PhD thesis, Université d'Aix-Marseille, **2012**.

[6] ISO 532 Method B – *Acoustics – Methods for calculating loudness level*. **1975**.

[7] J. B. Kruskal, M. Wish – *Multidimensional Scaling*. Sage Publications, Beverly Hills, **1975**.

[8] P. Legendre, L. Legendre – *Numerical Ecology*. Elsevier, Amsterdam, **1975**.

[9] O. Kirkeby, P. A. Nelson – *Digital filter design for inversion problems in sound reproduction.* Journal of the Audio Engineering Society, **47**(7/8), pp. 583-595, **1999**.