SUBJECTIVE EVALUATION OF THE SOUND QUALITY IN CARS BY THE AURALISATION TECHNIQUE

Angelo Farina (1), Emanuele Ugolotti (2)

(1) Dipartimento di Ingegneria Industriale, Università di Parma, Via delle Scienze - 43100 PARMA
tel. +39 521 905854 - fax +39 521 905705 –
E-MAIL: farina@pcfarina.eng.unipr.it – HTTP://pcfarina.eng.unipr.it

(2) ASK Automotive Industries, via Fratelli Cervi n. 79, 42100 Reggio Emilia
tel. +39 0522 388311 - fax. 0522 388499 - E-MAIL: tec_ask@xmail.ittc.it

0 INTRODUCTION

The evaluation of car sound systems is usually done both by means of objective measurements (frequency response, distortion, etc.) and by listening tests. The latter are particularly long and difficult, because it is necessary that the subject seats on each car and listens to a pre-defined music sample, played though the sound system. Furthermore, it is difficult to take into account the noise due to the engine and tyres, because it would be required to conduct the tests with the car running on a test track or inside a specially equipped laboratory. Actually the large number of subjective tests are made with the engine not running.

By employing the auralization technique, it is possible to prepare sound samples for making comparative subjective tests of the sound system of different cars: the sound tracks used for the subjective tests are not recorded inside the car compartments: instead they are reconstructed by convolving the original signal (a music sample taken from a commercial CD) with the binaural impulse responses previously measured for each channel of the sound system, and adding the car's noise, also synthesised on the basis of experimental measurement of the average noise spectrum.

The new technique is very fast to implement, does not require expensive instruments or tools, and makes it possible to conduct the listening tests everywhere requiring simply a notebook computer: this way a reasonable number of significant results were collected in a very little time, with minimum cost, and with the certainty that the results are not biased by the knowledge of the car's maker or by non-acoustical effects due to the furniture of the car or to other confort-related topics.

The details of the experimental technique and of the signal processing employed for the auralisation were already published, (Farina [1]), and here they are recalled only briefly. The main goal of this paper is to present the first subjective results, and the analysis of them which enables to perfect the goals of the research in course.

Although a detailed analysis of subjective results usually requires advanced statistical tests (Herman [2]), only basic data managing schemes were employed here: nevertheless, some very evident correlations were found, from which proper design criteria can be deducted, both for the sound system and for the compartment treatment.

Another very important point covered by this paper is the selection of the members of the judging panel: a preliminary listening test, conducted on a larger number of volunteers, revealed that there is a certain number of subjects who give inconsistent responses. If these subjects are included in any test, their "random" responses certainly degrade the overall coherence of the data, and often the spread can be so high that no evident correlation is found among the responses. The selection of the listeners revealed to be a very promising technique, for transforming the subjective test technique from an "approximate" diagnostic tool to an exact measurement.

1 MEASUREMENTS

2.1 Measurement of the background noise

In each of the 9 cars under tests, preliminary measurements of the interior noise, at various speeds, were conducted. The tests were made on an highway, at the three speeds of 90, 120 and 140 km/h.

A Bruel & Kjaer microphone type 4165 was mounted on a torso simulator, placed on the seat at the side of the driver. It was connected, through a B&K type 2231 Sound Level Meter, to a DAT recorder SONY DTC-790. A preliminary 94 dB, 1kHz calibration signal was recorded on each tape. A 20-minutes sound sample was recorded for each car, at each speed.

At the laboratory, the DAT recordings were played back and analysed through a B&K type 2133 real-time analyzer, and the 1/3 octave spectra were stored to disk and converted into a spreadsheet.

Figure 1 shows some of the measured spectra at the speed of 90 km/h.

The background noise recordings were not used directly for mixing with the auralised signals: instead, they were used as shaping filters for the creation of an artificial background noise, having the required spectrum, as it will be explained at chapter 3.

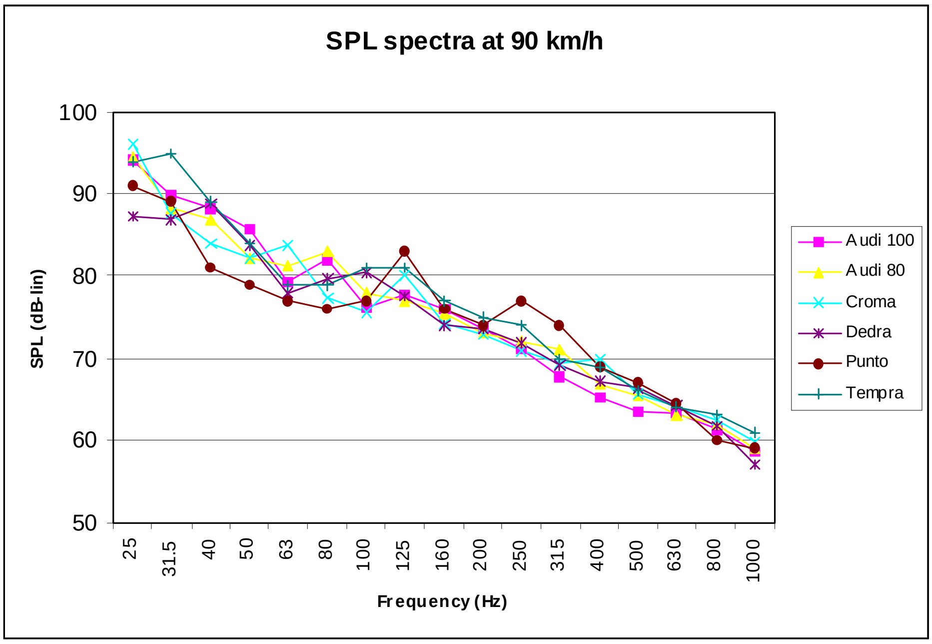

Fig. 1 – Background noise at 90 km/h

2.2 Measurement of the system's impulse responses

For a given position of the listener, 4 impulse responses (IR) have to be measured, as it is depicted in fig. 2: from each channel to each ear of the listener's head.

A further variable is the fact that some cars are equipped with a 4-way system, although the two rear channels are usually simply a copy of the two frontal ones. This required anyway two sets of measurements on such cars, one with only the frontal speakers, and the second with the complete system inserted.

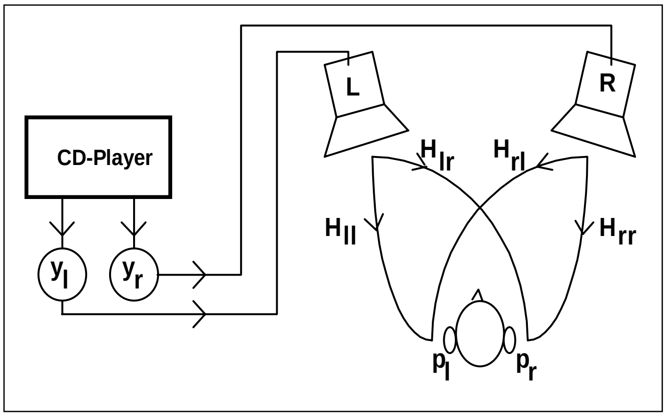

Fig. 2 – Scheme of the loudspeakers-ears transfer functions

The IR measurement was made employing a software MLS generator, and a software deconvolver for recovering the IR from a recording of the microphone signal, both running at a sampling frequency of 44.1 kHz. This system is being presented on a separate paper (Farina [3]). The signal, coming from the output of a 16-bit sound board incorporated in a notebook PC, was fed to the sound system by means of an electromagnetic coupler, inserted in the cassette player of each car (SONY CPA-4). This coupler was found to introduce an uneven frequency response, but it was easy to equalise the measurement results through a proper inverse filter, removing this effect.

A binaural dummy head was used for recording the signals (Sennheiser MKE2002), placed at the driver's position, and the microphone signals, properly pre-amplified through an home-made pre, were sampled through the line-in port of the notebook PC. Fig. 3 shows the dummy head inside a car compartment.

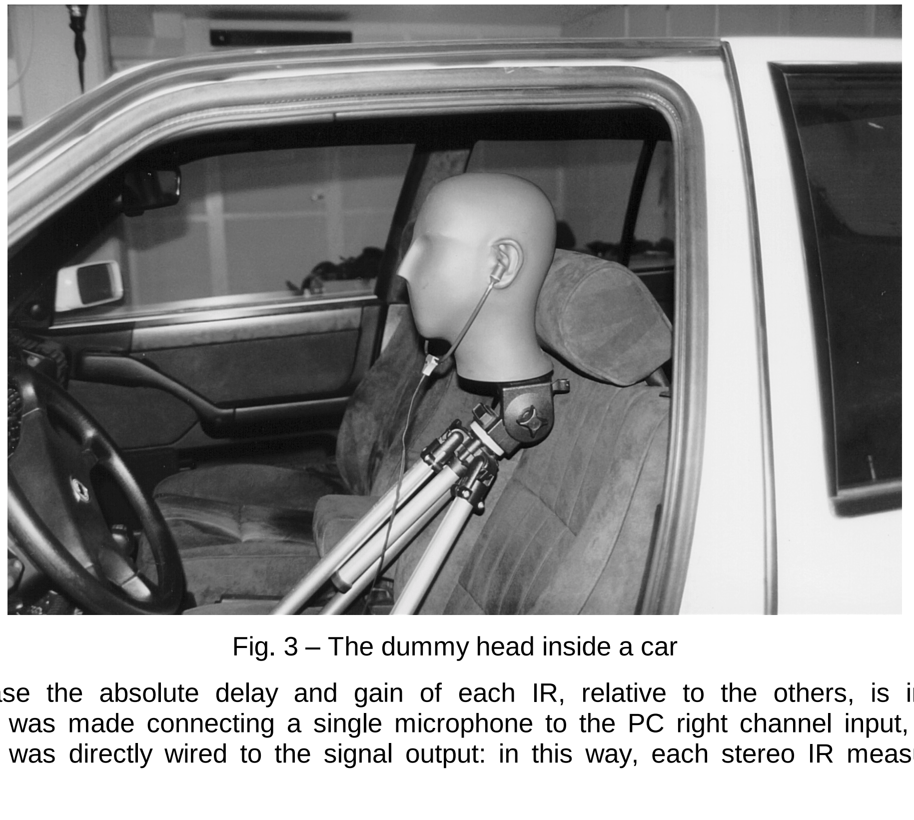

Fig. 3 – The dummy head inside a car

As in this case the absolute delay and gain of each IR, relative to the others, is important, the measurement was made connecting a single microphone to the PC right channel input, while the left channel input was directly wired to the signal output: in this way, each stereo IR measured contains always the same electric loopback signal on the left channel, with maximum amplitude and constant delay, and on the right channel the measured IR, with proper delay and relative amplitude. After stripping away the left channel information, the 4 measured IRs were packed into two stereo (binaural) IRs, and saved in .WAV format.

Fig. 4 shows the binaural IR of the left and right channels of a car.

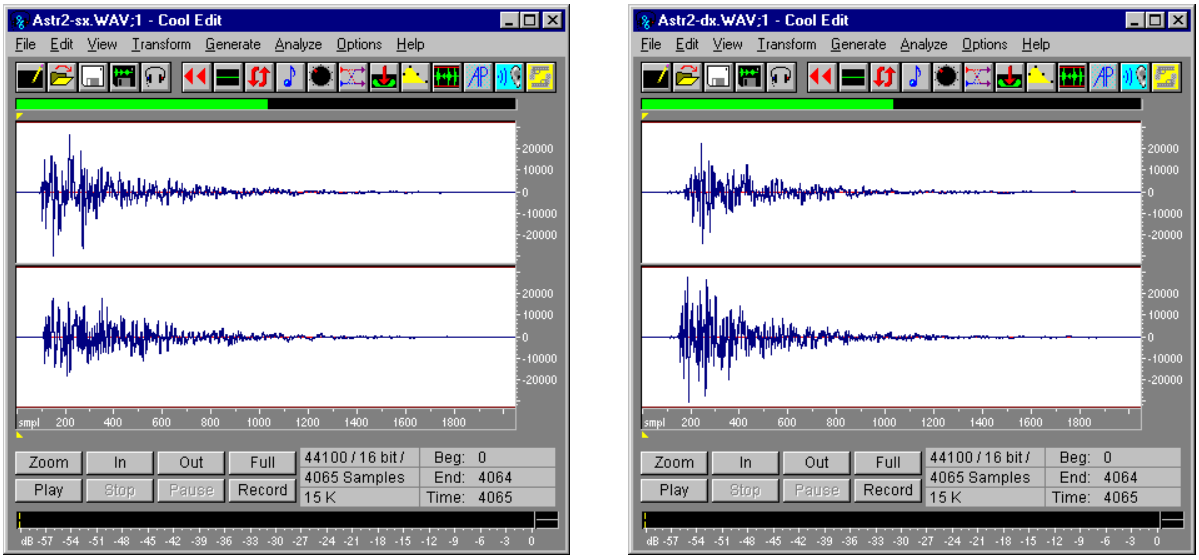

Fig. 4 – Binaural impulse responses of a car audio system (Astra)

2 SIGNAL PROCESSING

The signal processing is made of three different steps: convolution, noise superposition and presentation to the subjects (playback).

3.1 Auralization of the sound field

The first step is accomplished making use of a new software convolver (also presented in Farina [3]), which allows for the simultaneous convolution of a stereo original signal with two separate stereo IRs: so the whole process is fast and reliable, and the result can immediately be listened, or saved in a new .WAV file.

The original signals were samples of various kinds of music, digitally transferred from commercial CDs to the hard disk. Typically, the sound samples chosen for the tests were long between 30 s and 1 minute.

3.2 Background noise simulation

After the convolution was done, an equivalent background noise was generated, making use of the standard features of the sound editing program CoolEdit (also used for the other audio tasks already explained): we started with the generation of a “spatial-stereo” brown noise, and then we applied a proper frequency filtering, until the calculated 1/3 octave spectrum approached within +/- 1 dB the measured one, at the 90 km/h speed.

At this point the convolved signal was mixed with the background noise, taking into account the overall signal amplitude, in such a way that the absolute Equivalent Sound Pressure Level of the music at the ears of the subject was adjusted to 90 dB-lin, whilst the background noise was perceived with the same SPL as measured inside the car. This level adjustment revealed to be the most time consuming and delicate point of the whole signal processing.

This was also due to the fact that, listening to the reconstructed signals, it seemed that the background noise was too high compared with the music: the capabilities of our brain to concentrate only on the music when driving at a car, and neglecting the environmental noise, make so that anyone remembers his listening experiences on real cars with much lower noise than the reality. For discovering this point, and assuring the proper level calibration, some DAT recordings were made while playing a CD with the car running on the road: listening at such DAT recordings, the same anomalously high background noise is audible, while it was not perceived during the recording.This fact is one of the weak points of the new auralisation technique, because it causes a systematic overestimate of the subjective effect of the background noise, compared to the subjective experience reported when driving a real car.

3.3 Playback system

To present the reconstructed sound signals to the subjects, two reproduction systems were employed: the first is loudspeaker-based, the second makes use of headphones. Each of them needs to be properly equalised, for making it not influent on the results: for the headphones the task is accomplished simply by convolving the signal of each ear with a proper equalising FIR filter, whilst the loudspeaker reproduction system requires also a cross-talk cancellation scheme, for avoiding that the signal coming from the left speaker arrives also on the right ear and vice-versa.

The details of these inverse equalisations were already presented in other papers (Farina [1], Farina [3]), and so they are not explained here.

3.4 Subjective testing system

For automating the process of playing the sound samples and expressing the subjective judgements for each of them, a new software tool was developed. It is a .WAV player, equipped with a graphical interface for collecting the responses to a set of predefined questions. Both the list of .WAV files and of subjective questions are stored in ASCII files, so that the same program can be used for different subjective tests.

Each question is expressed as a couple of counter-posed terms (such as PLEASANT-UNPLEASANT), and the listener has to place a marker between them, on a 5-segment scale. So each response is represented by a numeric value, ranging between 1 (left term is more appropriated) to 5 (right term is more appropriated); 3 means that the response is in the middle.

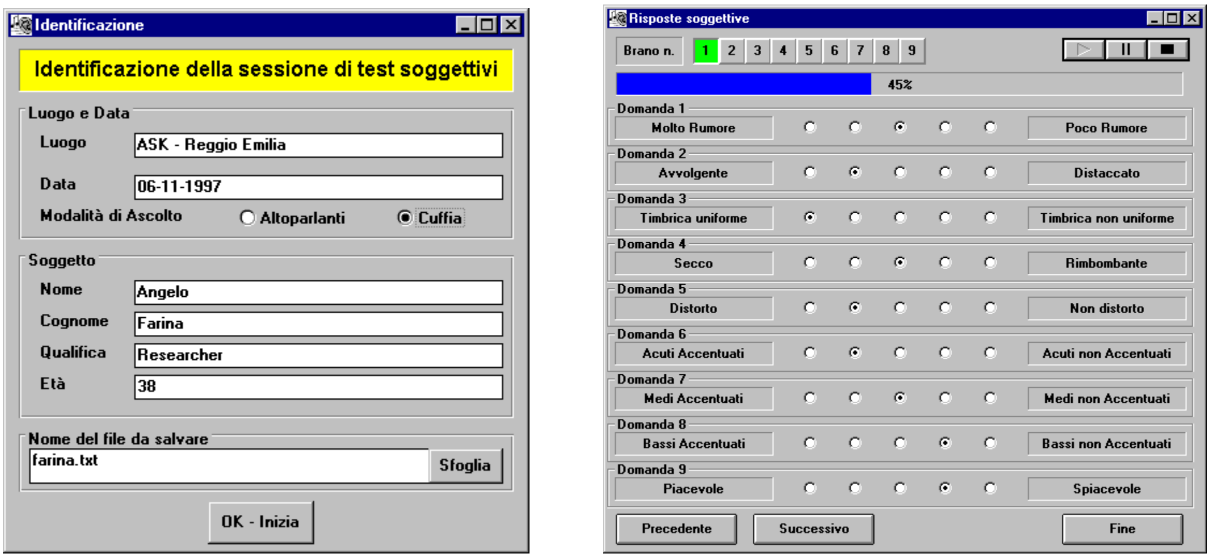

Fig. 5 – User's interface of the subjective test program

The user can change at any time the sample being played, pause the playback or start it again, come back and change some responses after listening at other samples, and he is left completely free to listen again at the sound samples, or to change the responses, until he is completely satisfied. Obviously he do not knows at what car each sound sample refers, nor he knows that the signals are artificially produced. Almost no one doubted that the signals were not naturally recorded inside running cars, but some complained about the noise “too loud”. Fig. 5 reports the user interface of the subjective testing program.

3 SUBJECTIVE TESTS

Two subjective tests were conducted: the first one had the goal of discriminating the listeners for the second one. Only those listeners who reached a good score in the preliminary test were admitted to the subsequent comparison test between cars.

4.1 Preliminary test

The first test was made presenting to 40 subjects 6 sound samples, which were heavily processed through software manipulation. The listener had to respond properly, making it clear that he was able of understanding the questions and of locating the artificial effects added to the signal.

For the preliminary test, these sound samples were presented

SOUND1	original, not filtered
SOUND2	mixed down from stereo to mono
SOUND3	low-pass filter, 6dB/oct at 2000 Hz
SOUND4	high-pass filter, 10 dB/oct at 500 Hz
SOUND5	distortion (4% THD)
SOUND6	copy of SOUND2 for consistency test

The questions were the following:

Q1	Distorted	Not Distorted
Q2	Treble enhanced	Treble reduced
Q3	Bass enhanced	Bass reduced
Q4	Stereophonic	Monophonic

It is clear that the “true” matrix of responses, expected from an ideal, sharp-eared listener, is the following:

	Sound1	Sound2	Sound3	Sound4	Sound5	Sound6
Q1	5	5	5	5	1	5
Q2	3	3	5	3	3	3
Q3	3	3	3	5	3	3
Q4	1	5	1	1	1	5

A global score for each subject can be obtained summing the deviation of each response from its ideal value. Fig. 6 reports the statistical analysis of the scores obtained from 40 subjects. It can be seen that the distribution is not perfectly Gaussian: instead, a sort of three-modal distribution has been found. This means that there is a small group of high-quality listeners, a second, larger group of average listeners, and a medium group of terribly bad listeners, who certainly have to be excluded from the test.

The average score was 23.5. For selecting only the good listeners among the others, an acceptance maximum score of 20 was selected. As it can be seen in fig. 6/right, which reports the individual scores, only 13 of 40 subjects were below this limit, and thus only these were employed for the subsequent comparative test.

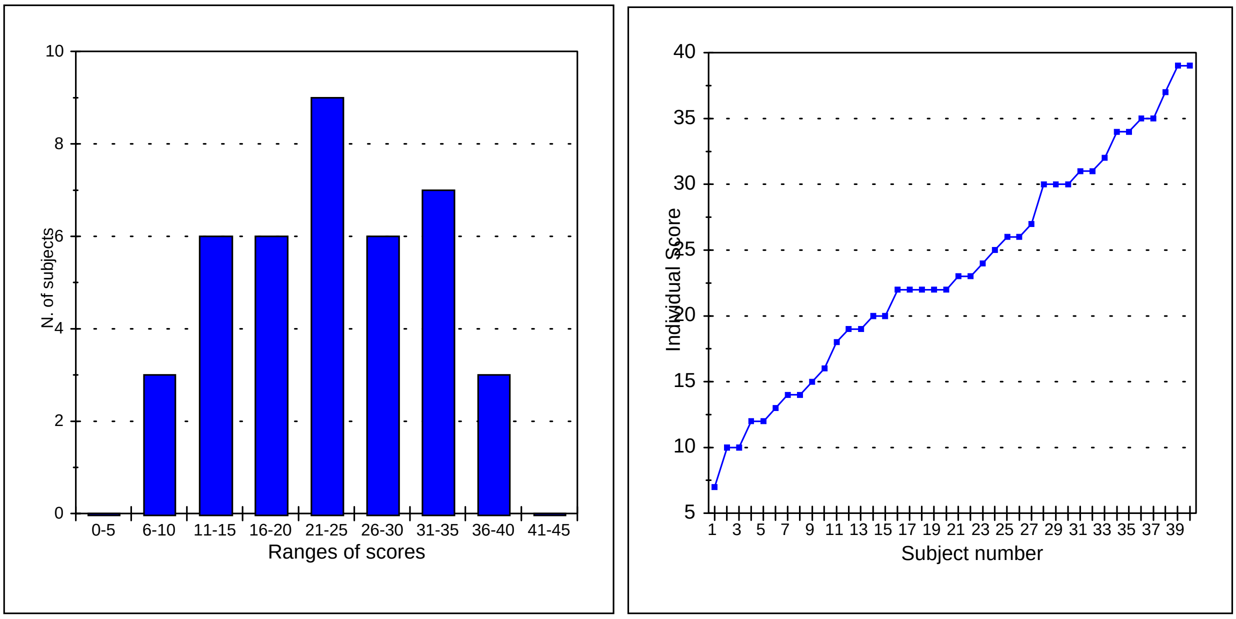

Fig. 6 – Statistical analysis of the preliminary subjective scores

4.2 Comparative test between cars

9 cars were employed for this comparative test. The sound system was always the original one as delivered by the manfacturer.

So the subjective test involved the comparison of 9 sound samples, one for each car. A different music piece was used in this case, of shorter length (30s), to avoid confusion with the preliminary tests, and to reduce the time required for completing the questionnaire.

9 questions were posed to the subjects, as reported (translated in English…) here:

Q1	Much Noise	Little Noise
Q2	Enveloping	Detached
Q3	Uniform timber	Not uniform timber
Q4	Dry	Reverberating
Q5	Distorted	Not distorted
Q6	Treble enhanced	Treble not enhanced
Q7	Medium enhanced	Medium not enhanced
Q8	Bass enhanced	Bass not enhanced
Q9	Pleasant	Unpleasant

In this case each of the 13 subjects produced a matrix of 81 judgements. This large amount of data was processed for finding the most evident correlations. The following table reports the average score of each car to the 9 questions, ranked following the score obtained for the question #9; remember that the lower the score, the most pleasant is the judgement:

Car	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Audi 100 B	3.31	3.00	2.69	2.69	3.92	3.15	2.85	2.85	2.62
Audi 80 B	2.92	3.08	2.69	2.15	3.54	2.92	2.54	3.46	3.00
BMW 735	3.77	4.00	3.00	1.92	3.31	2.69	2.69	3.92	3.31
Citroen Evasion	2.77	2.54	2.69	3.69	3.15	3.23	3.23	2.31	3.38
Opel Astra SW B	2.69	3.00	3.08	2.69	2.92	3.54	3.00	3.00	3.62
Fiat Croma B	2.15	3.54	3.38	2.85	3.00	3.38	2.92	3.46	3.77
VW Passat B	2.31	4.15	3.08	2.77	3.08	2.62	2.38	4.08	3.85
Dedra B	2.15	3.92	3.31	1.92	2.85	3.08	2.85	4.23	4.15
Punto B	2.08	3.69	3.85	2.62	2.31	3.92	2.69	4.31	4.46

Looking at the table, it can be observed how the overall pleasantness ranking is matched with the other subjective factors.

For making clearer these correlations, it is possible to make plots of the interdependence of pairs of subjective responses. Fig. 7 reports the diagram, which relates these two subjective parameters, and it is clear how the noise level is clearly negatively correlated with the acoustic quality. A linear trend is evident for 8 of the 9 cars, and the only exception is the BMW-735, which has very little noise, but nevertheless does not have a proportionally high quality judgement. Probably for this car other subjective parameters are more significant than background noise in explaining the overall quality score.

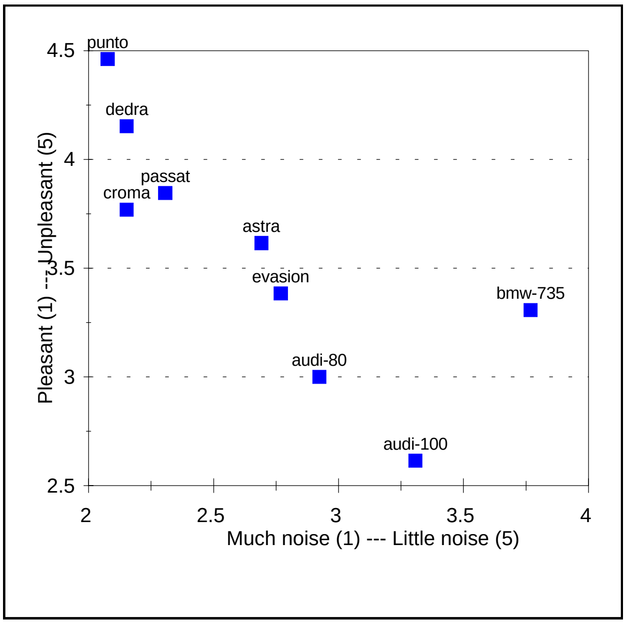

Fig. 7 – Correlation between average responses at questions #1 and #9.

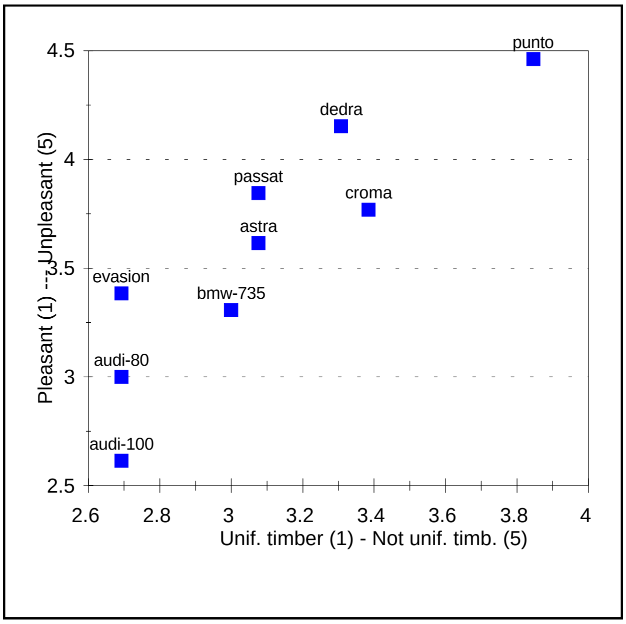

Fig. 8 – Correlation between average responses at questions #3 and #9.

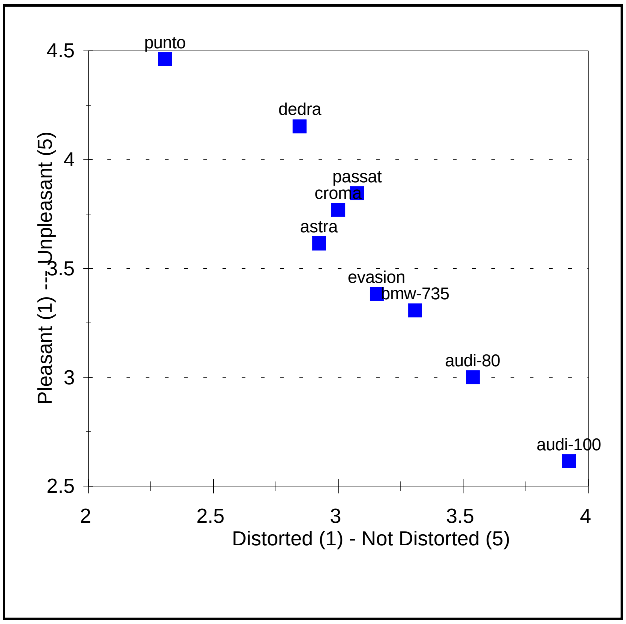

Fig. 9 – Correlation between average responses at questions #5 and #9.

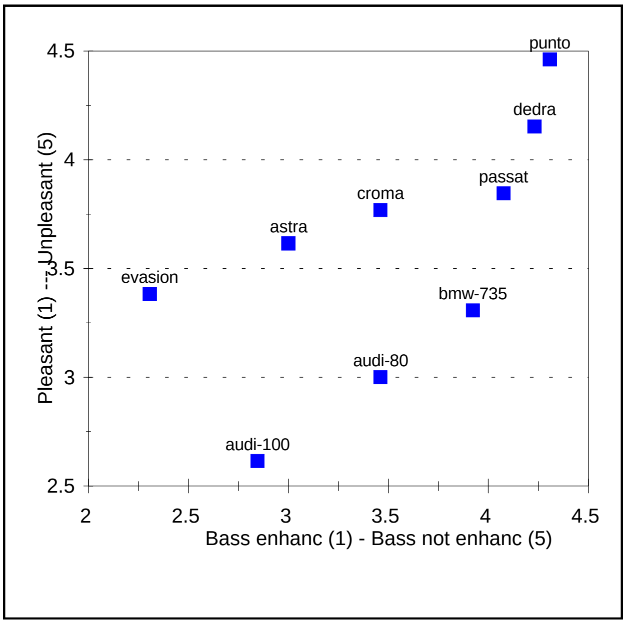

Fig. 10 – Correlation between average responses at questions #8 and #9.

Other interdependecies are shown on figg. 8, 9, and 10: they reports the other subjective responses which resulted well correlated with the global preference. The other subjective responses, instead, gave poor correlation with it.

Fig. 8 shows how the timbre uniformity is positively correlated with preference, although a certain spread is evident. Fig. 9 demonstrates how the distortion adversely affects the preference, with a quite good linearity of the plot, and fig. 10 relates the presence of bass frequencies with a general improvement of the preference, although in that case the spread is quite high. In this case a two-dimensional explanation can be attempted: in fact, the three cars with lower noise (Audi-100, Audi-80 and BMW-735) appear to lye on a lower line, while the other cars are grouped on an upper line: this means that a lot of bass is required particularly when the noise is high, for “covering” it.

Obviously at least these 4 subjective parameters interact each other in influencing the overall preference: the evaluation of the mutual effect of them is possible only with an advanced statistical technique, the factor analysis (Herman [2]), which in this case was not applied yet.

Nevertheless, the above information is enough for deducting some interesting conclusion, and for piloting the prosecution of the research.

4 CONCLUSION

A preliminary analysis of the subjective results shows that the auralization technique makes it possible to exploit the difference between the cars, as the other factors potentially influent on the subjective results are maintained absolutely constant. Furthermore, the total time required for conducting the experiment is largely reduced in comparison with the traditional technique based on direct binaural recordings. The last advantage is the possibility to evaluate directly any modification to the sound system or to the car compartment, by digital filtering of the measured impulse responses, for defining the most preferred characteristics of it.

The subjective results are consistent with already known design criteria, which demands for low noise inside the car compartments, for a reasonably flat frequency response, for absence of distortion and for wide extension of the frequency response toward the low frequencies. Although these requirements are substantially obvious, they confirm that the subjective method, applied to a selected panel of subjects, is a quantitative tool which gives sensitive results, also if the number of subjects is low.

The research will prosecute increasing the panel of subjects, and performing an exhaustive statistical analysis of the results. Furthermore, also the number of cars will be increased, with the goal of finding the “best car system”: making use of the new *virtual acoustics* technique presented in the parallel paper (Farina [5]), it will be possible to recreate such a best system inside any car.

In a subsequent phase, a correlation between subjective and objective parameters will be attempted, with the goal of defining proper design criteria for optimizing the subjectively perceived sound quality of car audio systems.

In a third phase, a numerical simulation of the sound field inside the car compartment will be attempted, following the guidelines given in Granier [4]: in this way it will be possible to state the acoustical quality of different design options, before any prototype of the system is built.

5 REFERENCES

[1] A. Farina, E. Ugolotti, ‘Subjective comparison of different car audio systems by the auralization technique’ , *Pre-prints of the 103rd AES Convention*, New York, 26-29 September 1997.

[2] Herman H.N., ‘Modern Factor Analysis’, *University of Chicago Press*, Chicago, 1968.

[3] A. Farina, F. Righini, ‘Software implementation of an MLS analyzer, with tools for convolution, auralization and inverse filtering’, *Pre-prints of the 103rd AES Convention*, New York, 26-29 September 1997.

[4] E. Granier, M. Kleiner et al., ‘Experimental auralization of car audio installations’ *Journal of the Audio Engineering Society*, vol. 44, n. 10, 1996 October, pp.835-849.

[5] A. Farina, E. Ugolotti, ‘Spatial equalisation of sound systems in cars by digital inverse filtering’*, Proc. of Reproduced Sound 13*, Windermere (GB) 23-26 October 1997.