Subjective comparison of different car audio systems by the auralization technique

Angelo Farina (*), Emanuele Ugolotti (**)

(*) 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

(**) 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

Abstract

This paper reports the results of a subjective evaluation experiment, based on the collection of questionnaires compiled by volunteers after listening to sound fields reconstructed by the auralization technique. Each synthetic sound field included both the car's interior noise and the transfer function of the sound system coupled with the passenger's compartment.

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 seat on each car and listen 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 synthesized 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 main goal of this paper is to detail the technique, regarding both the measurements inside cars and the auralization system employed for the listening tests: a comprehensive statistical analysis of the subjective judgements will be presented elsewhere [1].

In the following chapters the measurements made inside 9 car compartments are first described. Then the auralization/reproduction system is described, and a special subjective test computer program, developed for automating the collection of the subjective responses, is presented.

Some preliminary subjective results are then presented: although the number of subjects is still too low for making a definitive ranking of the 9 cars under tests, these results are already significant, showing that the new technique has great benefits against the traditional in-car tests.

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 analyzed 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, 2 and 3 show some of the measured spectra at the three speeds.

The background noise recordings were not used directly for mixing with the auralized 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.

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. 4: 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.

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 [2]. 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, as shown in fig. 5, but it was easy to equalize 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. 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 contained 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. 6 shows the binaural IR of the left and right channels of a car.

3. Auralization of the sound field

The auralization is made of two different steps: convolution and noise superposition.

The first step is accomplished making use of a new software convolver (also presented in [2]), 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, making use of the “Digital Domain” freeware program (by M. Overtoom) and a Toshiba CD-ROM unit. In this way, no re-quantization noise was added to the original sound, as it happens if the analog output of the CD player was sampled again with the audio board. Typically, the sound samples chosen for the tests were long between 30 s and 1 minute.

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 auralization process. 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: this fact is due to the capabilities of our brain to concentrate only on the music when driving at a car, and neglecting the environmental noise, so anyone remember 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 auralization 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.

Before the listening tests, anyway, it was necessary to equalize the auralized signals, for compensating the uneven frequency response of the reproduction system employed for the listening tests. This point will be explained in more detail in the next chapter.

4. 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 equalized, 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 equalizing 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.

Let we describe first the headphones equalization: the chosen headphones were placed on the Sennheiser dummy head, and an IR measurement was made for each ear (fortunately they resulted almost perfectly equal). Fig. 7 shows the average impulse response and frequency response of these headphones.

The inverse filter was created by the Morjopuolos least-squares technique through a proper software module [2], and it was convolved with the already prepared test samples of the 9 cars. In such a way, not only the frequency response of the headphones was compensated, but also the phase response, and the response of the microphones mounted on the Sennheiser dummy head too: this makes the headphone reproduction system completely blind. Fig. 8 shows the impulse response of the inverse filter and its frequency response, and fig. 9 reports the result obtained convolving the headphone response with the inverse filter: the time domain signature is an almost perfect Dirac's delta, and the frequency response is reasonably flat.

In the case of the loudspeaker reproduction system, a larger number of variables are involved other than the loudspeaker frequency response: speaker placement, room reflections, movement of the listener's head, matching between the dummy head response and the human listener's head response. In this case it is much more difficult to assure that the signal reconstructed at the ear channel's entrances of the listener are exactly the wanted ones.

The cross-talk scheme is again described by fig. 4: in this case it is necessary to measure again 4 impulse responses in the listening room, and to use them as input for the cross-talk cancellation module, described in [2]. This produces other 4 IRs, which have to be applied to the signals before sending them to the loudspeakers.

Both the dummy head and the real listener head can be used for the measurement of the 4 listening room IRs. Although the cross-talk cancellation obtainable with the listener's head is more accurate, this has two disadvantages: the sample signals have to be processed again for each subject, and the pinna's functions which are removed from the measurement chain are those of the listener, not those of the dummy head. For these reasons, the cross-talk cancellation was implemented making use of the 4 dummy-head IRs: fig. 10 shows the measured IRs in the listening room, fig. 11 the inverse filters computed by the cross-talk cancellation software, and fig. 12 reports the result of the convolution of the first set with the second. As it is expected, the LL and RR transfer functions are almost perfect Dirac's delta functions, and the LR and RL transfer functions are zero.

5. 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.

The program start with a simple dialog box, where some information about the subject is asked. After this is completed, the main form appears: on the top, a series of buttons allows for the choice of the sound sample, and below the list of questions appears.

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. 13 reports the user interface of the subjective testing program.

6. Result of the subjective tests

Two subjective tests were conducted: the first one had the goal of discriminating the listeners for the second one, presenting to them 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.

Only those listeners who reached a good score in the preliminary test were admitted to the subsequent comparison test between cars.

6.1 Preliminary test

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. 14 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-end listeners, a second, larger group of average listeners, and a medium group of terribly bad listeners, who certainly have to be excluded by 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. 15, 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.

6.2 Comparative test between cars

9 cars were employed for this comparative test. They are reported in the following table:

N.	Manufacturer	Model
1	AUDI	80 SW Benz
2	BMW	735 Benz
3	FIAT	Croma TD
4	LANCIA	Dedra Benz
5	CITROEN	Evasion Benz
6	Wolkswagen	Passat SW
7	FIAT	Punto-S Benz
8	OPEL	ASTRA SW Benz
9	AUDI	100 SW Benz

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

9 questions were posed to the subjects, as reported 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. A complete statistical analysis of these data is beyond the scope of this paper, and will be presented very soon elsewhere [1]. Here simply the results of question #9 are commented.

The following table reports the average score of each car to the question #9: remember that the lower the score, the most pleasant is the judgement:

Description	Q9 average score
car-9 = audi-100	2.62
car-1 = audi-80	3.00
car-2 = BMW-735	3.31
car-5 = evasion	3.38
car-8 = astra	3.62
car-3 = croma	3.77
car-6 = passat	3.85
car-4 = dedra	4.15
car-7 = punto	4.46

It is expected that with a proper analysis of the responses to the other questions it will be possible to understand the reasons of these results. In particular, a multiple correlation analysis, supported by an analysis of variance, could show the relevance of the other subjective parameters on the overall quality judgement, and also exploit the correlation between them.

Here a simple correlation between the noise-related parameter (question #1) and the subjective pleasantness (question #9) has been attempted. Fig. 16 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.

7. 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 research will prosecute increasing the panel of subjects, and performing an exhaustive statistical analysis of the results.

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 [3]: in this way it will be possible to state the acoustical quality of different design options, before any prototype of the system is built.

8. References

[1] A. Farina, E. Ugolotti – "Subjective evaluation of the sound quality in cars by the auralisation technique" – Proc. of 4th International Conference and Exhibition Comfort in the automotive industry - Bologna (Italy) October 2-3, 1997.

[2] 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.

[3] E. Granier, M. Kleiner et al. – "Experimental auralization of car audio installations" – JAES vol. 44, n. 10, 1996 October, pp.835-849.

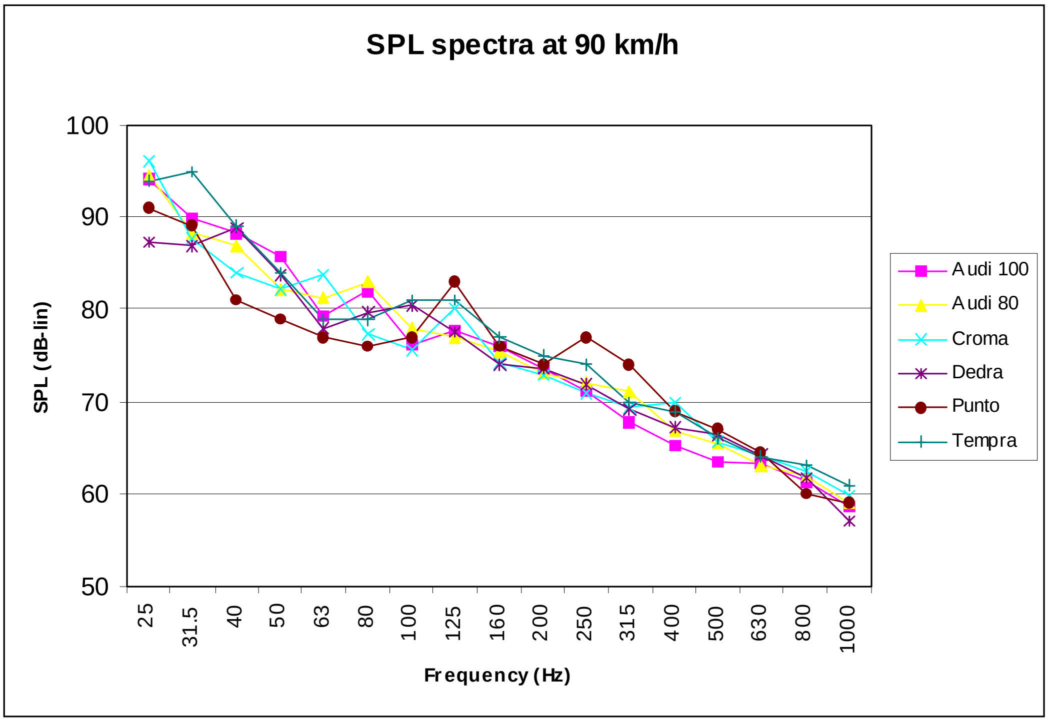

Fig. 1 – Background noise at 90 km/h

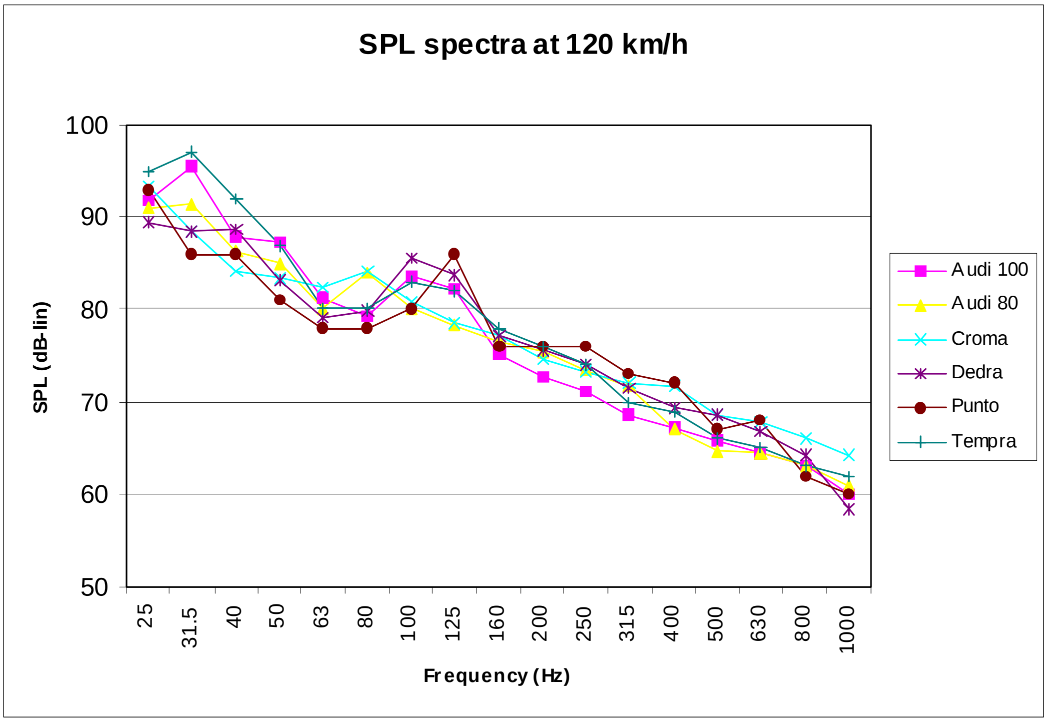

Fig. 2 – Background noise at 120 km/h

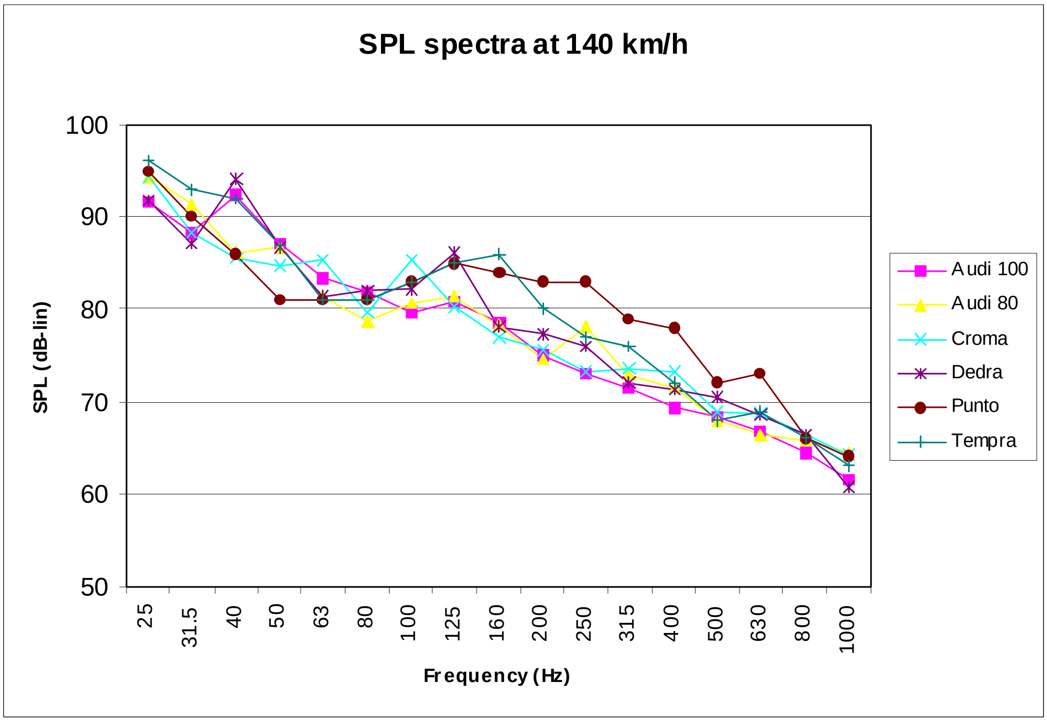

Fig. 3 – Background noise at 140 km/h

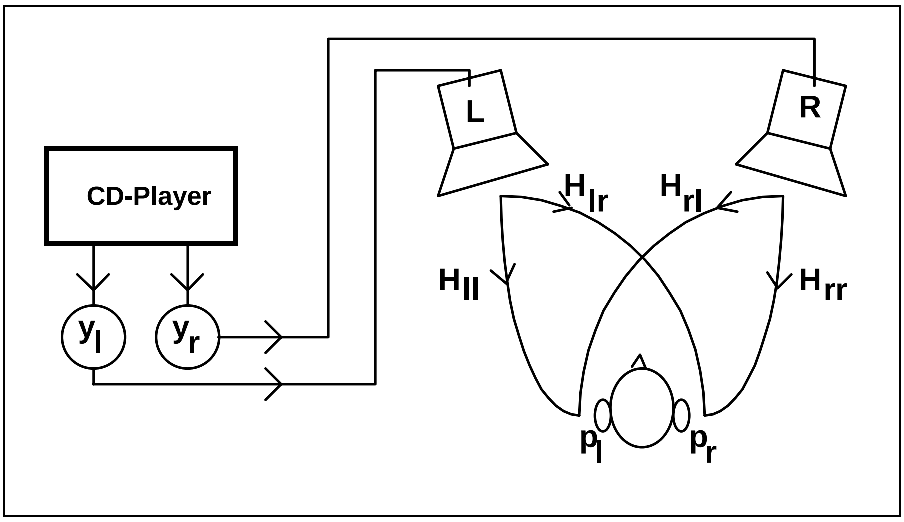

Fig. 4 – Scheme of the inter-aural crosstalk

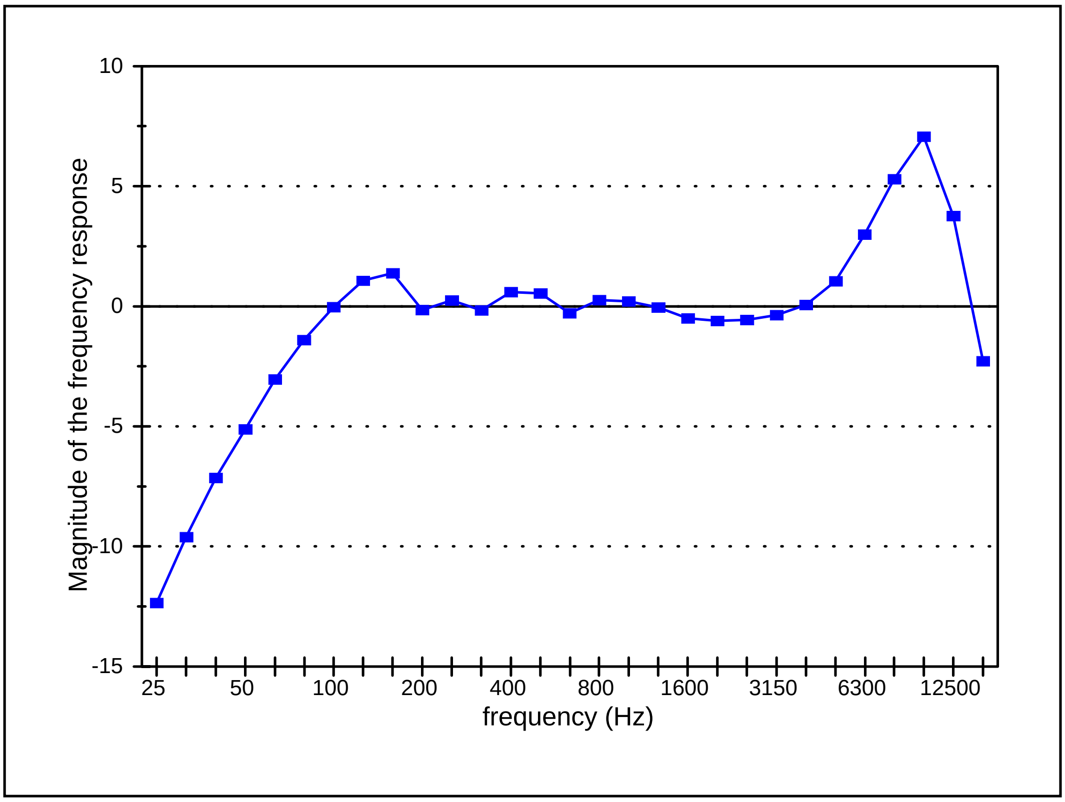

Fig. 5 – frequency response of the Sony Cassette Adaptor

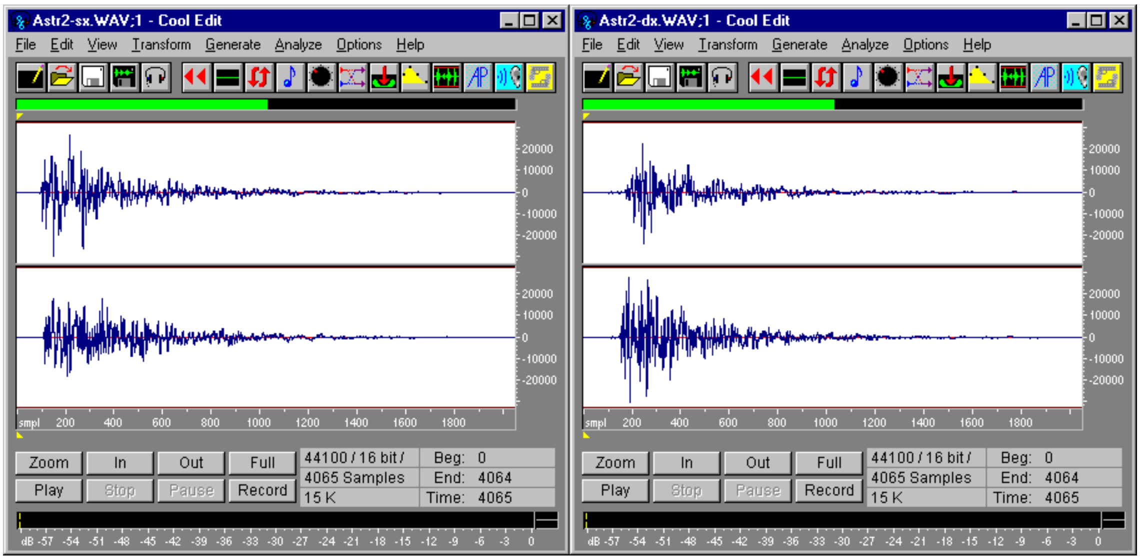

Fig. 6 – Binaural impulse responses of a car audio system (Astra)

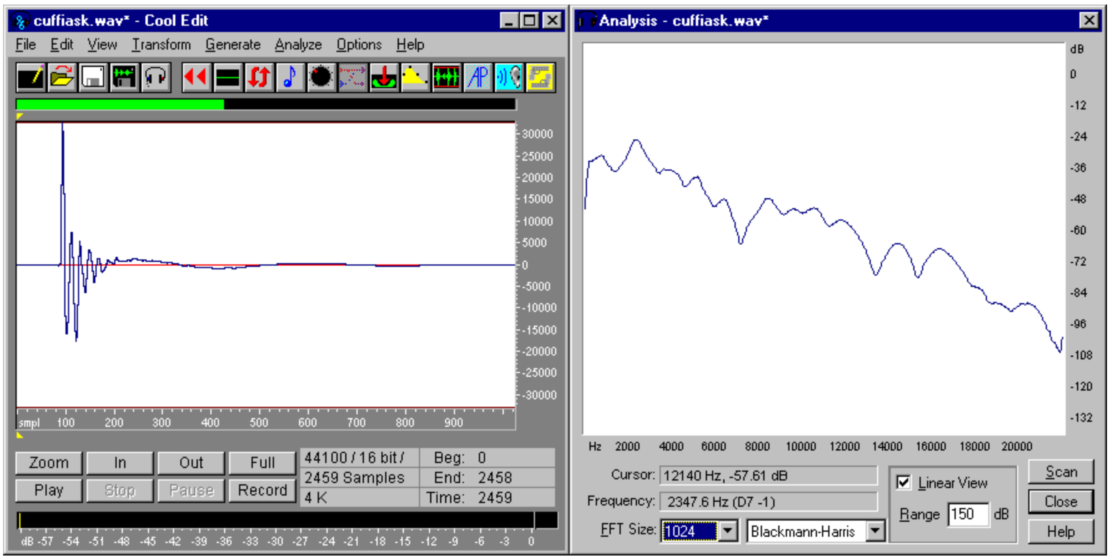

Fig. 7 – Impulse response and frequency response of the headphones

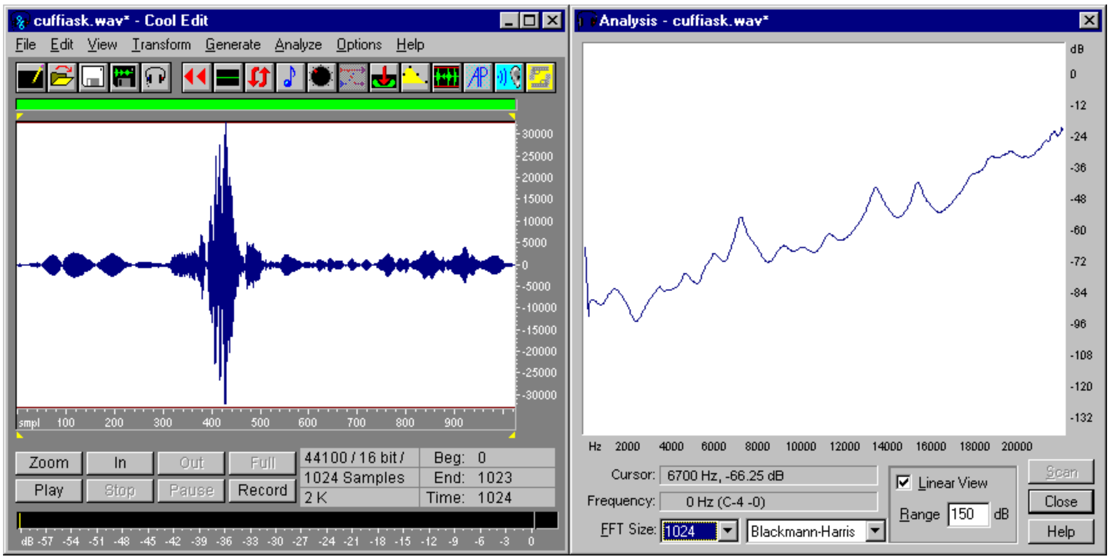

Fig. 8 – Impulse response and frequency response of the inverse filter

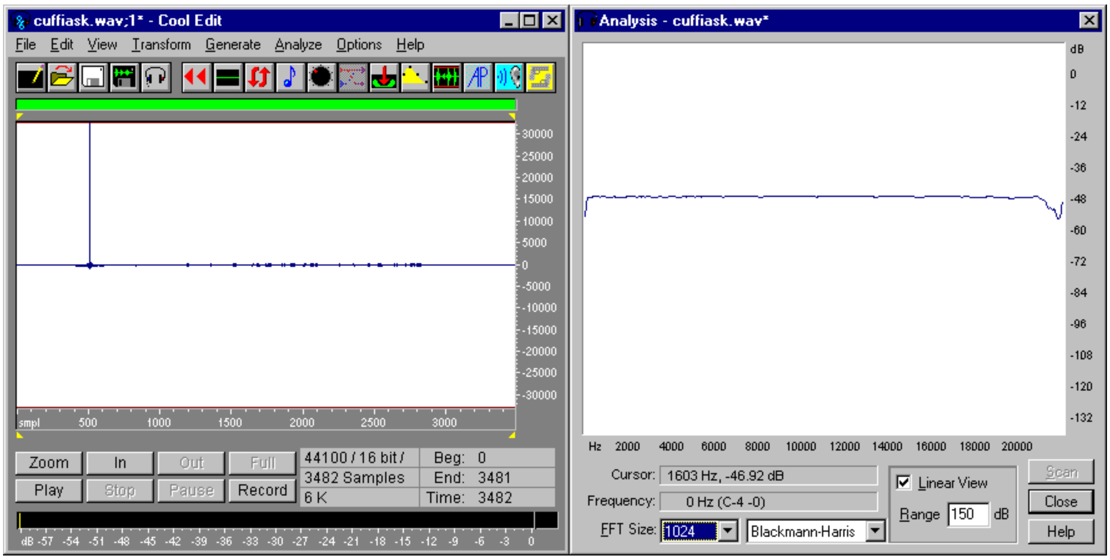

Fig. 9 – Impulse response and frequency response of the convolution of the headphone with the inverse filter

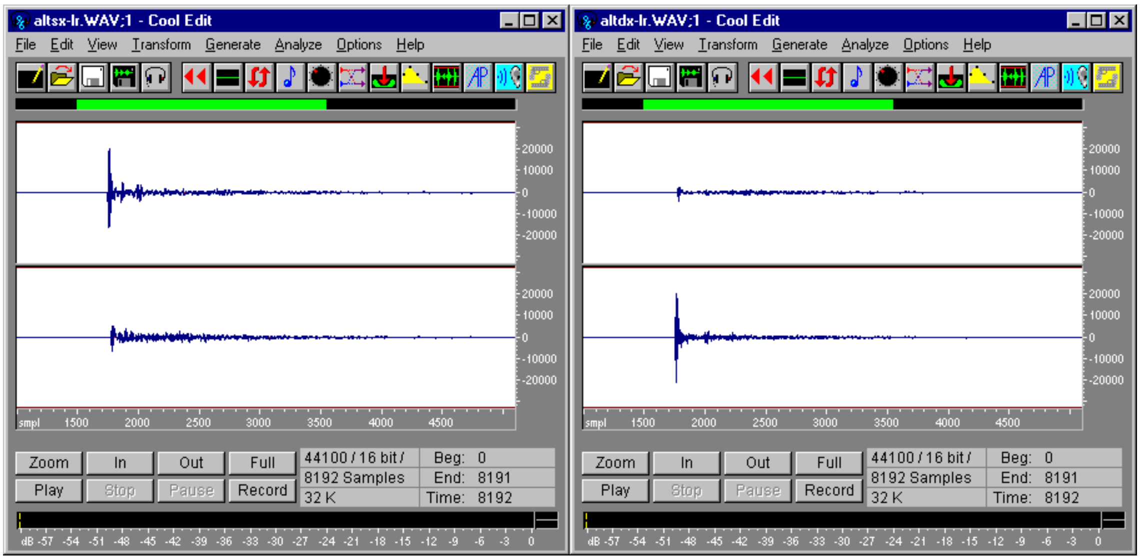

Fig. 10 – Binaural impulse responses of the ASK listening room

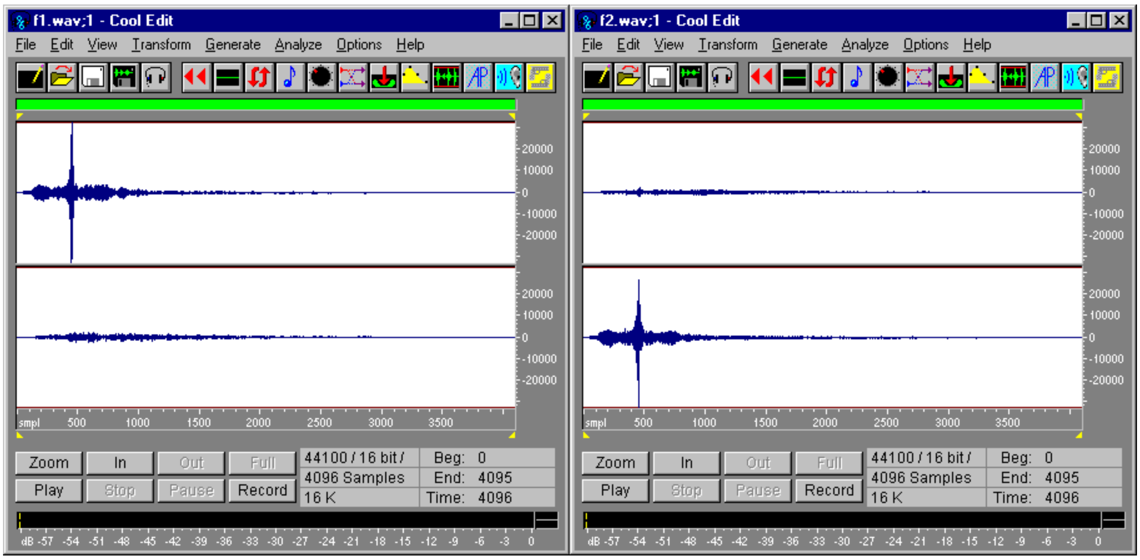

Fig. 11 – Inverse filters for cross-talk cancellation in the ASK listening room

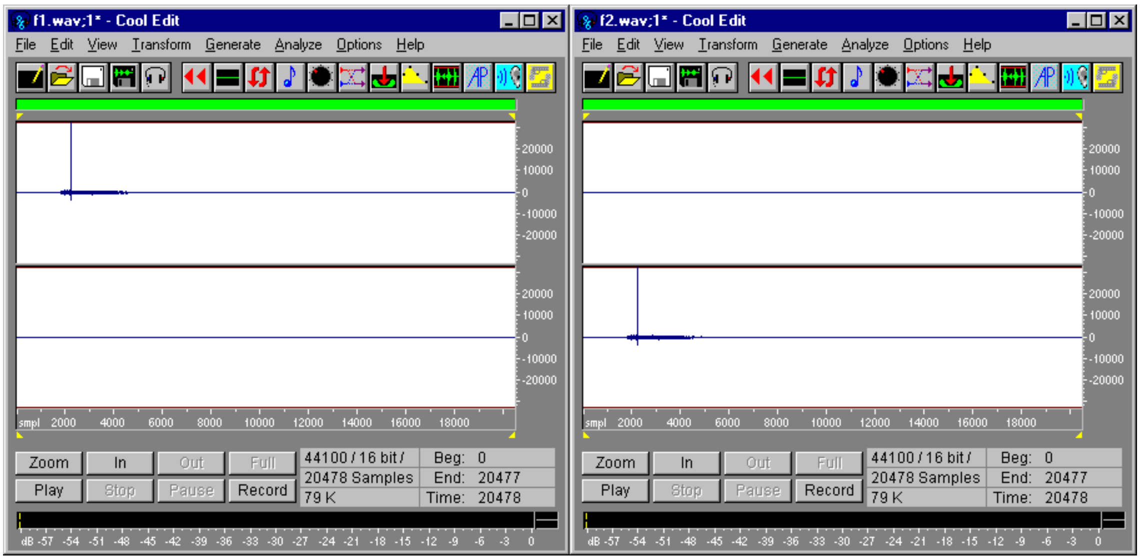

Fig. 12 – Effect of the convolution of the cross-talk filters with the listening room's responses

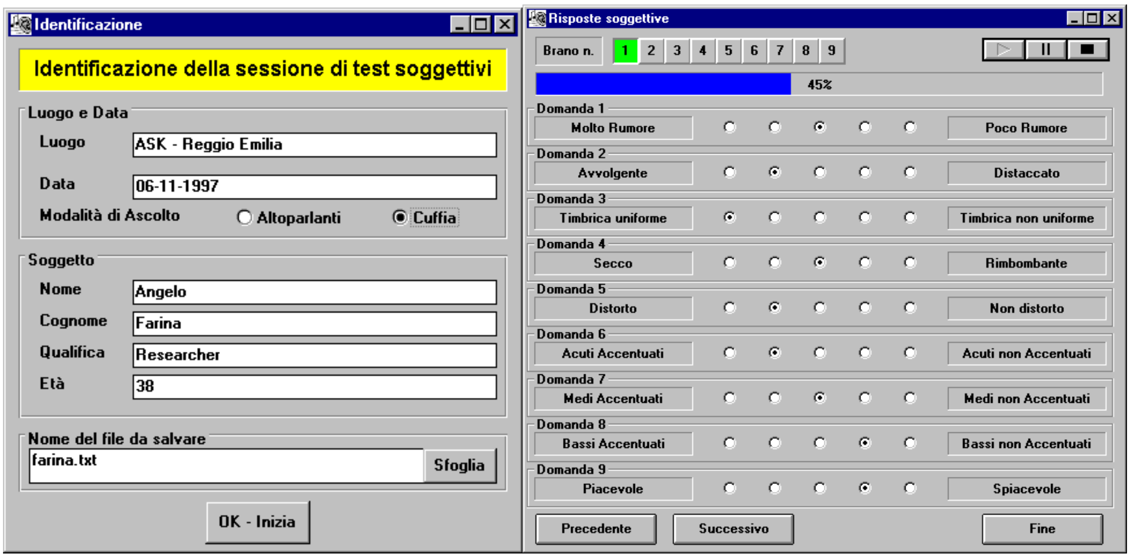

Fig. 13 – User's interface of the subjective test program

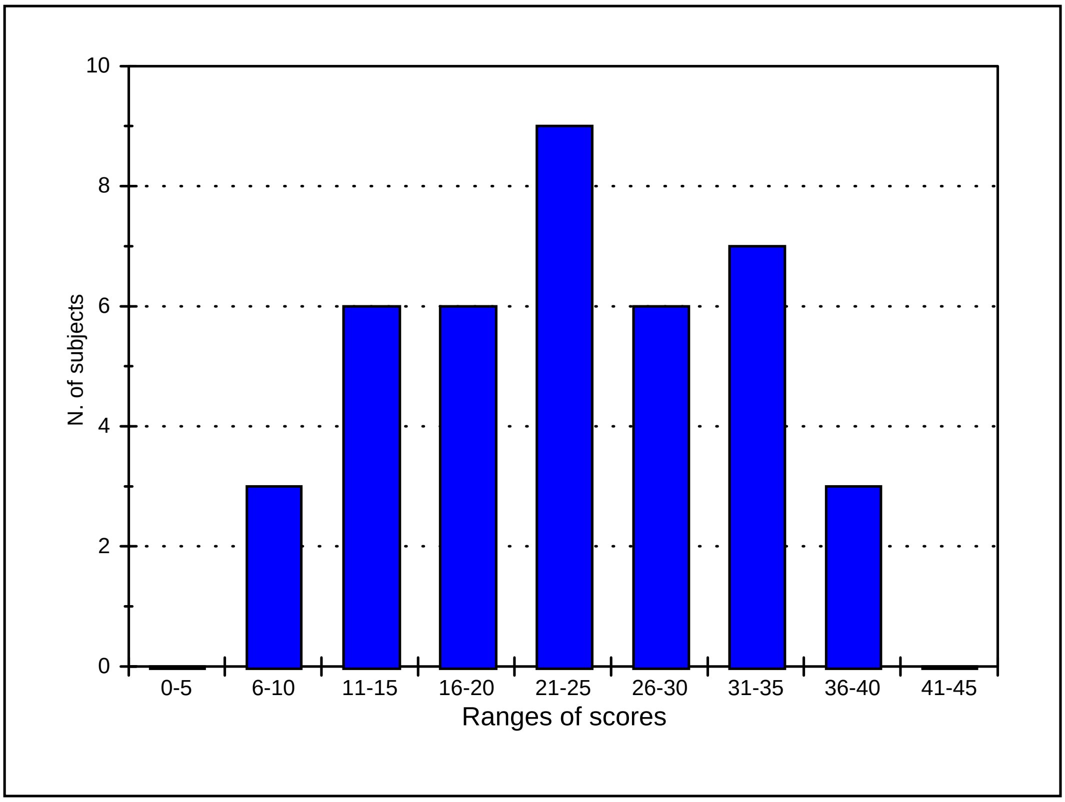

Fig. 14 – Statistical analysis of the preliminary subjective scores

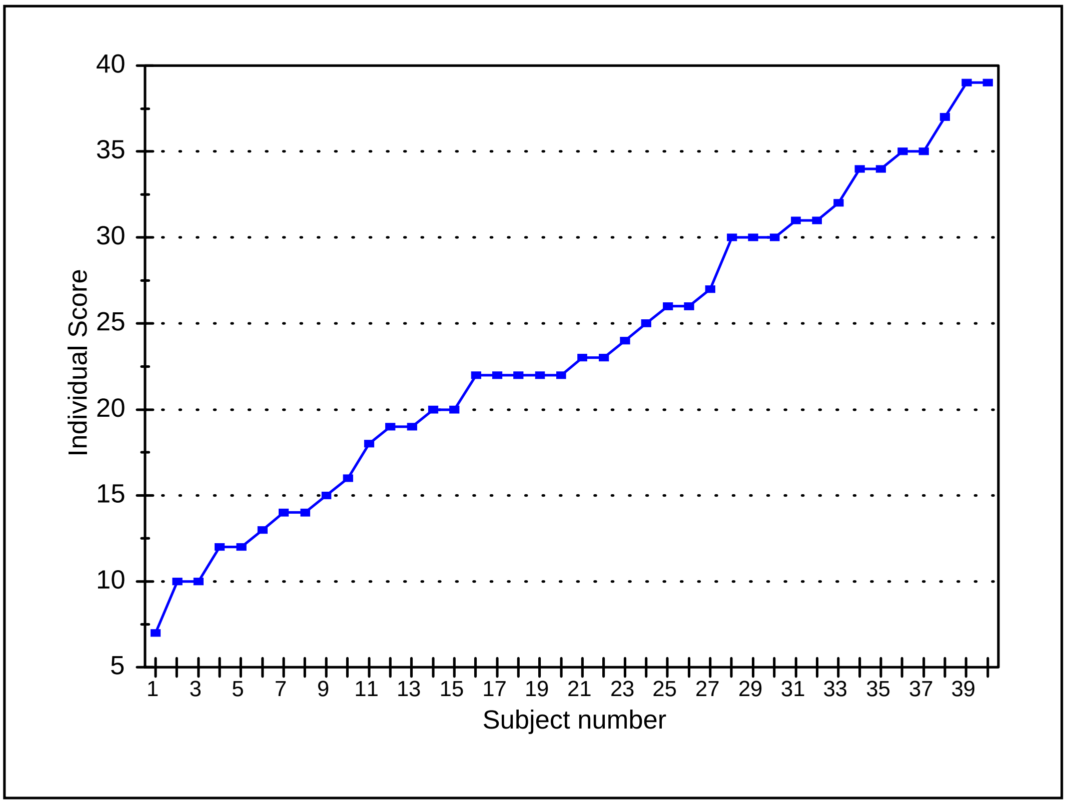

Fig. 15 – Individual subjective scores for the preliminary test

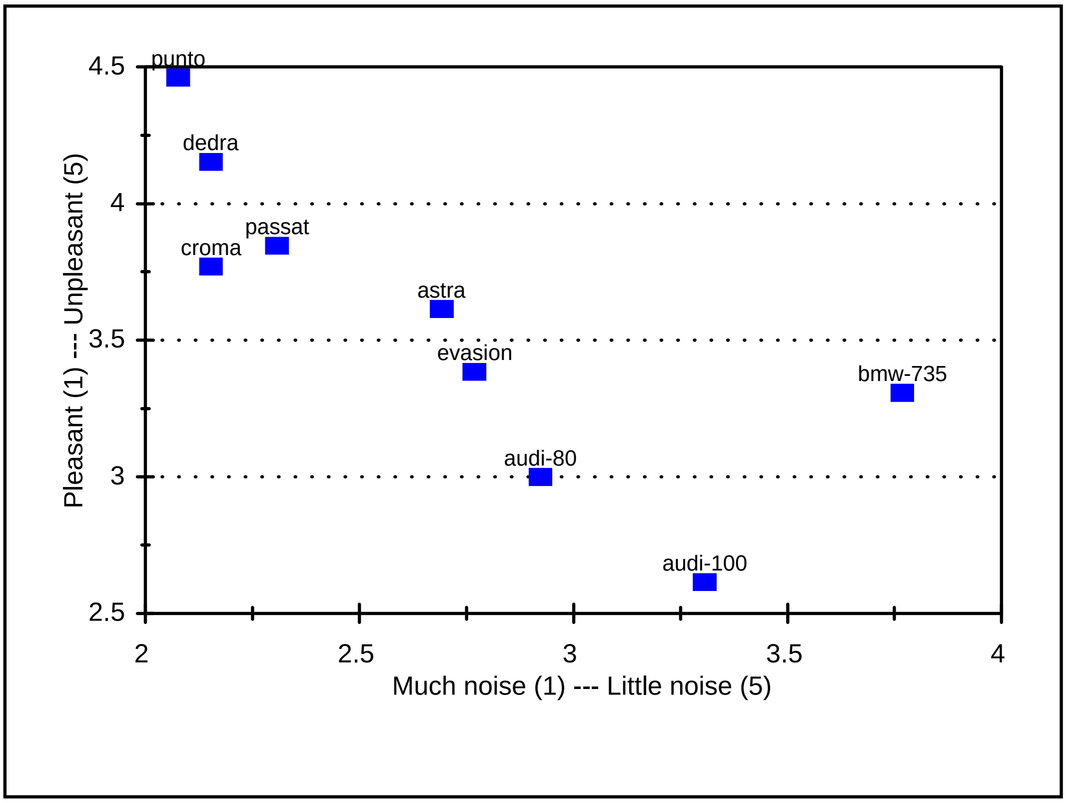

Fig. 16 – correlation between average responses at question #1 and #9.