Audio Engineering Society

Convention Paper

Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA

Construction of a Car Stereo Audio Quality Index

A. Azzali [1], A. Farina [1], G. Rovai [2], G. Boreanaz [3] and G. Irato [3]

[1] Industrial Engineering Dept., University of Parma,
via delle Scienze 181/A, 43100 Parma, Italy
Email address: farina@pcfarina.eng.unipr.it

[2] Fiat Auto – Infotainment Dept.
guido.rovai@fiat.com

[3] CRF – Vehicle Dept. – NVH
giovanni.boreanaz@crf.it, giorgio.irato@crf.it

ABSTRACT

A measurable index ("IQSB") quantifying perceived quality of car stereos has been developed, to forecast aural appreciation. Results of panel interviews and listening tests (in a special "auralisation room") have been correlated with the analysis of corresponding binaural recordings. Two outputs were obtained. First, a model of the subjectively most relevant features was identified, in terms of statistically significant "verbal descriptors". Second, a single-figure index was constructed, function of objective measurable quantities related with audio performance, and well correlating with the average verbal evaluation (both of "naïve" and "expert" listeners). This tool is of great importance for the automotive industry, because it allows for the direct quantification of the audio system performance, significant part of the perceived quality of the product.

1. INTRODUCTION

Car stereo systems, as original equipment, are becoming an increasingly relevant part of the customer's global comfort and satisfaction. Being able to have an early and quick quantification of the cost-performance ratio for available audio components is of utmost importance during the target setting and the validation phase. It is well known in fact that subjective tests for audio systems are unavoidable, but very expensive, since they require a lot of man power, and tricky, since untrained listeners can provide invalid results. A measurable index such as the one described hereafter can be considered as a robust first, usable result in that direction. The outcome of panel interviews and listening tests (carried out in a special "auralisation room" at ASK Industries, RE, Italy) have been correlated with the analysis of corresponding binaural recordings. Two kinds of output were obtained.

- Firstly, a model of the subjectively most relevant features was identified, in terms of statistically significant "verbal descriptors".

- Secondly, a single-figure index was constructed (function of measurable quantities related with audio performance), well correlating with the average subjective evaluation (both of "naïve" and "expert" listeners). This is the kind of tool needed to evaluate in advance what the perceived quality of the tested car stereo system will be, once it reaches the market.

2. DATABASE CONSTRUCTION

The work, aimed to the statistical correlation between the objective audio performance measurements and the subjective responses, started with the construction of two different and complementary sets of data (to be subsequently correlated):

- a set of binaural recordings of the audio signals produced inside ten different cars, comprehensive of "measurement" signals (for calibration, acoustic environment reproduction and analyses) and of a special selection of musical samples used for the listening tests

- a set of structured subjective responses, collected among population samples and jury panels, conceived to turn verbal and preference evaluations into numerical values to be statistically processed

2.1. Binaural recordings database

Up to now, the research about the acquisition systems has introduced several recording techniques with different features and levels of accuracy. For instance Binaural, Soundfield [1], Mark Poletti's [2] techniques are some among them. The choice of the best recording configuration depends on the aim of the research. For industrial purpose, binaural recording is the best compromise because it requires fewer resources and a quick measurement procedure, but it neglects velocity information of sound fields and reduces their vector components, acquiring only the time history of two dimensions of them. By analyzing the phase response of the two recorded channels, it is possible to obtain a two dimensional model of the sound field inside the cockpit. The presented system is based on binaural recording, according with the requirements of the automotive industry, that typically refers to this standard. Thus, a series of audio tracks and test signals was recorded with a B&K 4100D binaural dummy head for further activities of the research (listening test and performance analysis). A part of the research was focused on the standardization of the measuring procedure.

2.1.1. Measuring procedure

As it is known, reaching a stable and significant harmonic curve is one of the main difficulties during car audio characterization, due to strong dependence of the phase on position, mainly in the high-frequency region. So a novel measuring procedure has been developed to obtain a reliable and repeatable measure of the response inside the cockpit. A great part of this procedure was followed also for recording the tracks needed for subjective tests and auralisations inside a special listening room. The proposed standard has the following features.

Sound source

According to Industrial requirements, where time to market is one of the main aims, the fastest way for stimulating the audio system is a standard CD player. This kind of choice reduces the time needed for the preparation of the measure.

Test signals and musical tracks

A set of audio tracks and test signals, normalized to the same level, were stored in a CD. The audio tracks, needed for the listening test, were chosen in a list of well known and characterizing musical pieces. This list was proposed to a broad panel of listeners, that chose the most significant ones for testing audio performance. The choice of musical tracks was carried out looking to sound quality and spectral and spatial features of the recording.

The large experience of ASK Industries and University of Parma in measurement techniques [3],[4],[8] was employed to select the test signals for characterization of car audio system performance. As it will be shown hereafter, this test signals were used in analyzing impulse responses inside the car, and calculating objective parameters for constructing the quality index. The involved test signals are:

· *Pink Noise*: used for level calibration inside the car (see next chapter)

· *Logarithmic sweeps*: used for calculating distortion (THD), spatial parameters and the impulse response of the sound system [4]. The measurement of impulse responses based on this signal has a lot of advantages compared with the analogous MLS technique in terms of reliability and signal to noise ratio. Both mono and stereo sweeps were stored for stimulating the whole audio system and left/right channels separately.

· *Rotating signal*: a special signal was tested for characterization of localization features inside the cockpit. This novel approach in quality assessment is very important because it allows to estimate the performance of hi-end systems (like 5.1 ore 7.1 surround) and gives them a better score than standard stereo configuration. A better localization of sound sources in the recording can indeed be achieved using surround systems. So a quality index has to take into account also the localization features of a system.

Recording level

In order to investigate the dependence of the quality on the level, mainly in terms of distortion and spectral response, four different SPL's for recording of audio tracks and test signals were fixed: 80, 90, 95, 100 dB. Uncorrelated pink noise was uses to calibrate the audio system to this level. The further research, as it will be explained, shows how the recording made at 80 dB SPL is sufficiently representative, in statistical terms, of the system at the other levels. Moreover 80 dB resulted to be the most pleasant level for listening (without background noise). The perceived quality and optimal level in real driving conditions has also been investigated.

Dummy Head Position inside the car

The proposed standard takes into account also the driver's position. In fact the need for reaching a reliable and stable measurement requires that the listener's position be fixed and univocally determined both for audio and test signal recording. The chosen position derives from a mean of the positions assumed by common people during driving. In this way the first measurement position was fixed. Both test signals and musical tracks were recorded leaving the dummy head in this configuration.

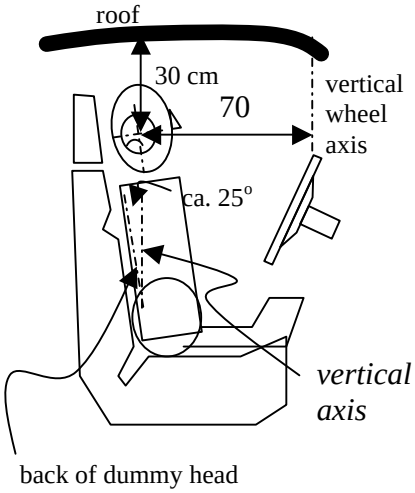

Figure 1.1: Vertical position of dummy head with torso

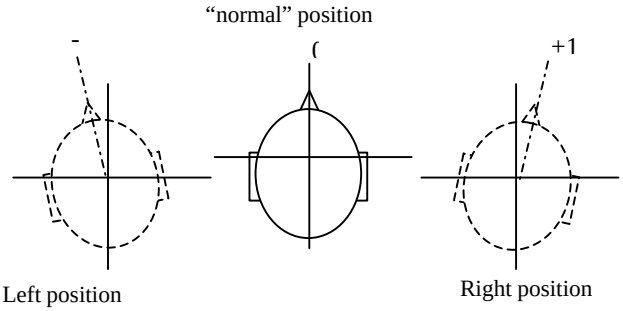

Figure 1.2: Rotation of dummy head

The dependence of the phase from position typically makes the measurement of the harmonic response unstable and not repeatable. So only one measuring position is not characterizing and sufficiently reliable for calculating real spectral responses. Several impulse responses of the system have to be measured in several measuring positions. According to real driving conditions, in which the driver moves the head performing imperceptible rotations, four measuring positions were fixed. The first has already been fixed. The others were obtained by rotating the dummy head of plus and minus 15 degrees away from the longitudinal axis, and elevating the head (in the frontal

position) of 5 cm. The CD containing the test signals was recorded in this four positions and four stereo impulse responses were obtained in post-processing. The spectral responses were averaged and a stable and reliable measure was obtained (see next chapter).

2.2. Subjective response database collection

The first thing that has to be taken into account is that the subjective database has been constructed by means of a dynamic process. That is, each step of the data collection was based on the result of the analyses carried out in the previous phase. For instance, expressions obtained with the first questionnaire were used to construct the questions contained in the second one. And so on. Some assumptions were inferred from previous works, too. An investigation methodology was prepared, structured in four consecutive steps, in order to:

- *identify the profile of a sample of the customers population and the terminology they use to describe what they intend for "quality of car stereo systems".*

This goal was achieved by means of a questionnaire consisting in a first section ('subject's profile') concerning personal data and experience with stereo systems (both domestic and automotive versions), and in a second section with a number of questions asking for subjective description of audio quality features ('audio quality profile'). The questionnaire was sent by e-mail or on paper to about 160 people; more than 100 of them returned the completed form. In the 'subject's profile' section, questions were asked about age, gender, qualification and profession. Questions were asked, too, about experience with domestic stereo systems (for instance, about the number and typology of owned stereo devices, about the average time spent in listening to them, about the amount of reading on the subject etc.) and, similarly, about specific experience with car stereo systems. In the 'audio quality profile' section, definitions of the main quality factors (not only aural features) for a car stereo set were asked. Then descriptions, opinions and keywords were requested, that could give a verbal representation of the interviewee's own idea of "quality" concerning the reproduction of sound inside the vehicle. The questions were structured in such a way as to permit linguistic and frequency analyses. They were basically "open" questions (that is, questions in which no selection within a predefined fixed set of choices is forced).

- *select the most "robust" terminology subset, that can be considered as common across the population sample, so that it can be used for "subjective" descriptions of acoustic events that can be unambiguously understood with the same meaning by (nearly) everybody.*

To this aim, the recurrent terms and locutions found with the first questionnaire were used to construct a second one (e-mailed to the same population sample, and 77 copies of which were returned, duly completed). It proposed questions (often "closed questions") in which the subject was mainly requested to choose among groups of items only the ones that sounded more "expressive" for defining quality features. In particular, verbal polarities between opposite descriptors were sought. Moreover, personal data were requested again, in order to get possible population subdivisions. In more detail, the subject was requested to:

>> evaluate the importance of each of 60 verbal expressions on a 3 level scale ("to be ignored", "slightly meaningful", "very meaningful").
>> add to each of them a list of possible synonyms and opposites.
>> try to group them under the categories found after analyzing the first questionnaire (asking to describe each category itself with a word or a locution), referring to "clusters" of sound quality features
>> score the relative importance of each of the above mentioned "sound quality categories" (by means of 7 grades summing up to 100)

- *prepare a listening test, based on "virtual auralisation" techniques, in which the selected "terminology core" can be used to link "meaningful" verbal descriptions with controlled acoustic environments.*

Two "jury panels" were used: 30 "naïve" subjects, plus 9 "experts" (with relevant experience in the acoustical or in the musical field). The assumption was made (and validated at the end of the test) that the population of "experts" could provide more "stable" results ("subject-independent"), so that a smaller number of them was sufficient. The listeners worked in the listening room at ASK by means of a specially developed SW interface

running on a PC (see paragraph 2.2) that allowed them both to listen to the different sound samples (and to compare them, if needed) and to evaluate their acoustical qualities on a "semantic differential" scale, based on the couples of opposite descriptors found with the second questionnaire. The scores were directly recorded on the database.
What they actually listened to, thank to the implemented auralisation method, were the almost exact reproductions of the acoustic fields as they were sampled inside the cars. In this way, direct comparison of the different aural environments (compartment + stereo system) was made possible, comparison that would have been almost impossible with real vehicles.

- *check the listeners' aural ability, in order to properly weight their individual contribution to the results of the listening test (results of "good listeners" have been assigned a higher weight than those of "poor listeners").*

The details of the four investigation steps are described in detail in the next paragraph.

3. SUBJECTIVE MODEL CONSTRUCTION

3.1. The preliminary questionnaires

First, the distribution of the answers to questions concerning personal data in the first questionnaire was statistically analysed, in order to monitor possible subdivisions in "clusters" of different behaviours among the observed population. In general terms, it seems that attitudes towards sound systems evaluation don't show significant variation in function of individual profiles, except for an aspect that, once demonstrated with further investigation, could prove to be of some importance for audio devices manufacturers and industrial end-users. It appears that when people show particular interest for the use of domestic stereo sets, they don't care so much about the quality of the automotive ones, and vice versa. In other words, the fans of domestic systems and those of automotive ones could be, after all, two distinguishable sub-populations of customers, with different tastes.
Other data about the studied population sample can be summarised as follows:

- the average number of domestic stereo sets possessed by the interviewees (comprehensive of portable and PC systems) is 3.4 (14% of this population owns more than 5 different sets), pointing towards the conclusion that widespread interest and familiarity with audio products could be expected.

- 63% of the interviewed people declares they "often" or "very often" listen to their stereo, so that familiarity with sound quality categories (at least at an experiential level) could be expected, too.

- about 60% reads the instruction manual (but this also means that about 40% doesn't read it…), less than 20% consults websites or reads magazines about the "audio" subject (most of the panel was contacted by e-mail and are supposed to share some basic technological experience): this seems to point towards the idea that some experience with stereo systems is quite common among the involved population section, but seldom with a significant degree of knowledge.

- about 50% spends between 1 and 3 hours a day driving a car, about 45% spend less than 1 hour, the rest more than 3 hours, and more than 80% "always" or "almost always" listens to the stereo when driving. Thus, it can be said that the "target customer" can be interested in audio entertainment in the car for less than 3 hours a day

- car stereo enthusiasts seem to show an attitude towards being more available than domestic stereo enthusiasts to spend money for high quality systems.

Then, the answers concerning verbalisation of perceived audio quality features were analysed. The various given verbal descriptions were grouped in seven "basic" categories, by means of the study of common meanings (the authors also established "heuristic" links between the words used by the members of the panel and physical acoustic features) and comparison with the results of previous works carried out at ASK and the University of Parma. The seven found categories were (the original in Italian language is given in brackets):

"fidelity" ("fedeltà"): capability of a system to reproduce sounds as similar as possible (to the human ear) to the sounds emitted by the original ("real") source

"cleanness" ("pulizia"): absence of unwanted disturbances in the sound signal, at any volume

"intensity" ("intensità"): capability of a system to reproduce loud sounds without disruptions of the signal and the subjective feeling of "excessive volume"

"location of sounds in space": capability of the system to give the impression that sounds recorded from different, well separated acoustic sources come from different locations in space (around the listener). For usual stereo systems this acoustic space is generally limited to an horizontal plan, that should be placed at the height of the listener's ears. After the analysis of subjective choices, this category was renamed as "origin of the sounds" ("provenienza dei suoni")

"diffusion [of sound] in space" ("diffusione spaziale"): capability of a stereo system to reproduce a sense of "immersion" in the acoustic field, that is the listener should feel the impression of being "surrounded" by an acoustic landscape coming from all directions

"sound character": this expression refers to the most subjective features of "colour" of the reproduced sound, the aesthetic properties that get closer to the individual tastes

"bass and treble tones": capability of a stereo system to correctly reproduce the extreme frequency regions (very low and very high audible frequencies), in a way matching as far as possible the individual tastes. . After the analysis of subjective choices, this category was renamed as "bass/treble balance" ("bilanciamento di bassi e alti")

A category was then added, because of its peculiar importance in audio reproduction, that didn't emerge from the analysis of the answers, that is "voice quality" ("qualità della voce"). This was probably due to the attitude of the interviewees towards thinking of the stereo set as a device mainly aimed to the reproduction of musical sounds, thing that can have introduced some bias in the results. As a matter of fact, a posteriori comparison between statistical models with or without this descriptor (included in the testing protocol) has shown that its inclusion is meaningful. Then, the analysis of the subjective expressions about synonyms and opposites gave the verbal extremes for each category, to be used for the semantic differential evaluations (for instance, the extremes for the "fidelity" category were "distorted" and "true", in Italian "distorto" and "fedele", that is an adjective etymologically linked with the word "fedeltà").

3.2. The listening test

3.2.1. Listening test facilities

The subjective listening test was carried out using a user friendly software interface, to evaluate the appreciations and scores for each set of verbal descriptors. The questionnaire was proposed in the ASK listening room, designed and developed with the aid of University of Parma. Hereafter this tools will be shown in detail.

Virtual listening room of ASK Industries – RE – Italy

A part of the research was spent to equip a realistic and high-performance listening room allowing to reproduce, with high-fidelity and transparency, audio tracks recorded in car compartments. In this activity the main goal of the listening room was to maintain in the virtual environment the invariance of quality judgments expressed inside the real cars. In fact, normally, the reproduction system doesn't have a truly flat response and distorts harmonic, spatial and dynamic behaviour of the recorded tracks. For instance, headsets are not appropriate reproduction systems, because the listening conditions are different from the real ones, where we have a frontal sound field and a sound arriving from all around. In headphone systems we have only a lateral fraction of the whole field, and in particular conditions the sound seems to come from inside the head, with a complete loss of the spatial information. We can achieve better results with headphone-auralisation techniques, but we cannot obtain frontal impression, and real spatial sensation. Moreover, not well-known psychological factors could interfere with subjective reactions. Several researchers dealt with this topic, and, by performing subjective test, demonstrated that a Cross-Talk Cancellation system (in the following CTC) formed by two loudspeakers better retrieves the features of the original sound sources than headphones and results are more correlated with the real listening. Starting from binaural standard recordings, an innovative reproduction system has been developed, in order to obtain better results from the two channels, overworking all the stored information. Placement of the loudspeakers in a particular configuration was

performed in order to improve the performance of standard CTC. In fact, traditional stereo dipole systems are not able to accurately reproduce the recording made inside the cockpit, because of several reasons: highly reflective surfaces (lateral and frontal windows) and the very small cockpit size make the two recorded signals from a binaural head strongly correlated to each other, with a loss of stereophony. The CTC works fine with highly stereophonic recordings, but it has some problems with the recording inside the car. The sound has also a relevant frontal component, that the CTC is not able to reproduce accurately. The developed system improves the performance of a standard stereo dipole system and allows to achieve a better performance, using eight channels, accurately processed.

Figure 2: Listening room and virtual reproduction system

An improvement of the stereophony and of the frontal fraction of the sound field has been achieved with digital filtering, loudspeaker placement and signal processing. A subjective test was performed to validate this system. The real placement of the loudspeakers will be made public in future. An equalization algorithm, loaded on a DSP platform, was adopted to correct the response. The filters are loaded on three DSP boards (Analog Device SHARC EZ-kit ADSP21161N).

The capacity of the room to leave unchanged the scores given during subjective investigation inside the car was investigated. So a blind subjective test, aimed to validate the effectiveness of virtual system, was performed. A questionnaire, composed by six questions, was proposed to a judging panel of ten trained listeners that work in automotive field. At last a Visual Basic software was developed to make the comparative test faster and automatic. The results of the test are proposed here in tab. 1 and 2:

Car	*1*	*2*	*3*
Spatiality	7,65	3,19	6,66
Clean treble	7,70	3,66	6,31
Clean voice	5,69	3,00	5,68
Clean and P. Bass	6,69	3,61	5,34
Pleasantness	7,31	2,75	5,99
Final Score	**6,89**	**3,44**	**5,88**

Table 1: Result of the test inside the listening room

Car	*1*	*2*	*3*
Spatiality	7,32	4,56	7,00
Clean treble	7,65	4,35	7,30
Clean voice	6,00	5,33	7,03
Clean and P. Bass	8,31	4,19	6,35
Pleasantness	7,32	4,00	7,30
Final Score	**7,34**	**4,60**	**6,86**

Table 2: Result of the test inside the cars

Subjective test

After complete validation of the facilities and of the base methodological tools, they were used to carry out the real listening test. The polarity scales obtained by means of the preliminary questionnaires were used to construct the software interface by means of which the listeners (the 30 naïve listeners and the 9 experts) were asked to:

- perform a preliminary “dummy” test (in this section the listener’s response data were not recorded) mainly aimed to get familiarity both with the test and the database of sound samples (reproduced at the same SPL until the last two sections of the test)
- compare a set of constant musical samples, reproduced (by means of the auralisation system) as they would sound in the interior of different cars, and score them on a first scale of “global pleasantness”
- repeat the comparison focusing on each of the eight defined categories making up the subjective model of perceived quality
- repeat the comparison concerning the categories of “intensity” and “fidelity” (that were more sensitive to this particular aspect), controlling at will the volume of the reproduction, the level of which was recorded together with the response data
- repeat the comparison concerning the global pleasantness, controlling at will the volume of the reproduction, the level of which was recorded together with the response data (this last section mainly served as

a check of the variations of perceived quality with volume changes)

The listeners could interact with the system interface by means of the PC mouse, to "jump" freely from the listening of a sample to another. This procedure allowed for what could be defined as an "implicit pair comparison" sequence, in which only the relevant comparisons were performed by the subjects. In this way, the sensory advantages of direct comparison were preserved, while the unpractical size of a complete, traditional pair comparison experimental design (which would have introduced severe decay in listeners' performance) was avoided. The score was assigned also by means of the mouse, operating a "virtual" slider on the screen that moved along the polarity scale. It was possible for the subjects to correct each score at will, if it happened that they changed their mind while getting more familiar with the "active listening" performance. The scores given to each specific descriptor were considered as the independent variables, the global evaluation as the dependent one (global evaluation = *f[descriptor 1, …, descriptor n]).* A regressive linear model was then constructed, connecting the specific and the overall subjective judgments, in which only the statistically significant components survived.

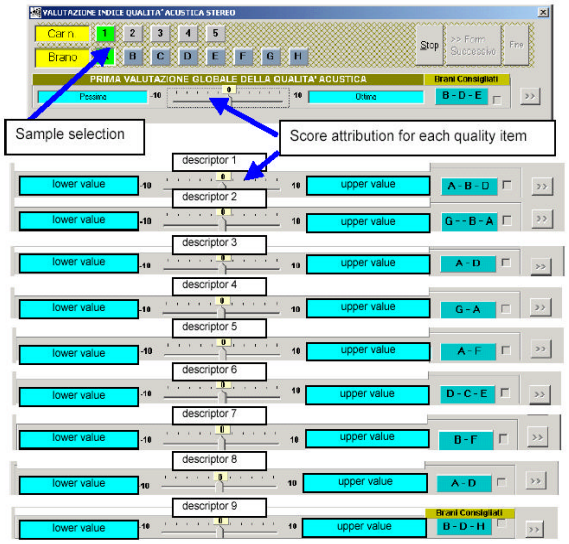

Figure 3: Visual Basic Software for test

As explained before, the answers of each subject were inserted in the database with a weighting factor, calculated on the basis of statistical coherence tests on the produced rankings, so that the answers of the most "clever" listeners were assigned a greater statistical weight (see the weight distribution in table 3). As it could be expected, the performance of the expert listeners proved to be better than the one of the "naïve" listeners (due to a smaller number of ranking inconsistencies).

subject	12	29	17	14	30	22	20	8	27	1
weight	0.0	0.4	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8

subject	24	19	16	3	6	9	107	21	15	13
weight	0.8	0.8	0.8	0.9	1.0	1.1	1.1	1.1	1.2	1.2

subject	105	18	7	11	23	100	26	10	28	25
weight	1.2	1.2	1.2	1.3	1.3	1.4	1.4	1.4	1.5	1.5

subject	106	4	2	108	102	103	101	104		
weight	1.7	1.7	1.7	1.9	2.0	2.1	2.2	2.2		

Table 3: weights assigned to each subject. The expert listeners have a coloured background, and are numbered starting from 100. The weights of the expert listeners is higher the naïve one almost in all cases.

Particularly the determination of the scores was carried out as addition of three indices. For each subject two cars were fixed and re-proposed during the two sessions of the test.

Index1: related with the level of consistence during evaluation of the descriptors on the two fixed cars. The index2 was used if the ranking wasn't respected. Instead index3 was used if it was.

$$Index2 = - S_{i=1,\dots 14} \{abs(vet^{i}_{1,1} - veti_{2,1}) + abs(vet^{i}_{1,2} - vet^{i}_{2,2}) \}$$

$$Index3 = - S_{i=1,\dots 14} \{abs(abs(vet^{i}_{1,1} - vet^{i}_{2,1}) - abs(vet^{i}_{1,2} - vet^{i}_{2,2})) + abs((vet^{i}_{1,1} + vet^{i}_{2,1}) - (vet^{i}_{1,2} + vet^{i}_{2,2}))\}$$

with $vet^{i}_{k,j}$ = evaluation of k=1,2 car during j=1,2 session for i=1, … 14 (8+variants) descriptors.

$$Weight = 3*index1+index2+index3$$

After the statistical analysis of the results of the listening tests, only three descriptors survived in the regressive model: fidelity, character, balance of bass and treble. As explained before, a fourth descriptor was added because of its special meaning, the "energy of the speech region", as representative of the "voice quality". As a consequence, the model had to be in the following form:

global evaluation =	const.	+
	a* "fidelity"	+
	b* "character"	+
	c * "bass/treble balance"	
	(+ d * "*speech region*")	

In the next picture the scoring of global evaluation obtained by the ten proposed cars are presented:

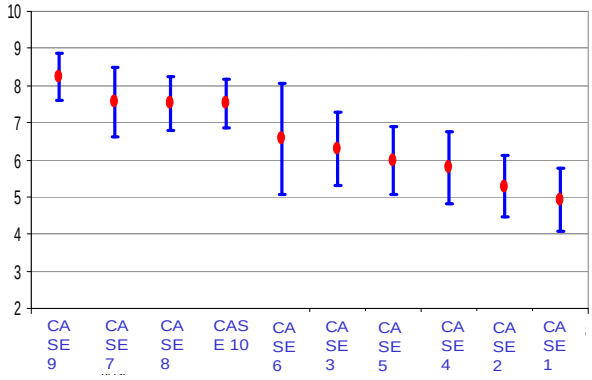

Figure 4: Global scores - "naïve" test

At this point the "subjective" representation of the perceptual model was completed, the global "like – dislike" general judgment on the aural performance of a car stereo system was "decomposed" into its main subjective factors. The next step was to connect them with well-correlated measurable parameters allowing for the construction of a prevision index capable of emulating an average human response.

4. OBJECTIVE ANALYSES

Many measurable acoustic, psycho-acoustic and stereo panning related parameters were evaluated for the digital samples of the test signals recorded in the ten cars. They were chosen to reflect in the physical world each of the eight found verbal descriptors. In practice, a large set of measurable variables was clustered under the label of a corresponding descriptor, according to "reasonable" knowledge based links, leaving to the subsequent statistical processing the goal to select those showing a really significant correlation with the subjective responses.

For instance, the subjective evaluation of "intensity" was related with parameters such as SPL, A-weighted SPL and 2 measures of psycho-acoustic loudness; the impression of "fidelity" was tested against harmonic distortion parameters and deviations from experimental "target" transfer functions; and so on. On the whole, 279 different (or "slightly different") acoustic parameters underwent the statistical selection process, based on correlation and significance analyses and feature selection / downsizing algorithms. In the following, some hints about the used parameters are given, grouped in five macro areas.

4.1. Spectral Parameters

Spectral behaviour of the car compartment and sound system was investigated using a novel approach to audio analysis: AQT Method. This method introduces a great improvement in quality assessment, because retrieves curves and parameters more correlated with subjective evaluation of acoustic pleasure. In fact AQT estimates the real perceived curve of human hearing, more sensitive to transients and peaks, analysing the dynamic behaviour of the system. On the contrary, classical FFT analysis retrieves information about steady state condition. Moreover musical signals are composed by transients, as well, and can be better represented by bursts. In the following the theory on which the method is based, is presented.

4.1.1. AQT ANALYSIS and AQT Tool: NEAR MUSICAL STIMULI AND AUDIO SYSTEM ACQUISITION

Characterization of the response inside a car is a difficult task because it is affected by a lot of barely known factors. The choice of the stimuli, of the microphone, of the position are only a few examples. A great step towards the complete characterization of hearing inside the car was made with the introduction of AQT Analysis. The acquisition process is complex, and human hearing system is more sensitive to transitory events because of masking, and Haas effect. In

summary, attack and release transients in sound events are far more relevant than the steady state information. It turns out that traditional methods show important limits in giving a complete characterization of the propagation in a car. Moreover, the musical signal is strongly non-stationary. If we equalize the response of the car in a static way, we cannot obtain a reasonable quality improvement.

AQT is a pre-existent method introduced by Liberatore [5]. The first version of this method was used for listening tests and as a graphic chart. AQT stimuli signals were played in the room used for the test and the responses were recorded and drawn in a chart. In this paper this method is extended to equalizer synthesis.

The true innovation introduced by AQT method is the stimulus signal. In order to measure the dynamic response of the system the stimulus is a train of sine bursts with variable frequency. This stimulus is more close to music, characterized by transitory events, and it allows to compute resonance frequencies. Then the response to AQT is close to the human hearing process, since we compute the values of response during attack transients for each frequency. Haas effect is accounted for keeping the duration of the burst at each frequency longer than human hearing system integration time. AQT analysis produces two parameters which provide a quantitative evaluation of these effects:

· Articulation: it estimates the speed of energetic recovery in an environment. Let's assume that we are playing a burst for a period of 200ms at a given frequency. After this, due to reflections, a certain time is needed before the energy extinguishes. Then a tail is associated to each frequency, the length of which depends on the absorbing properties of the materials inside the car. The frequency with a long tail response will be strongly affected by masking effects. On the other side shorter tails and higher articulations corresponds to higher dynamic.

· Dynamic harmonic magnitude response: it represents the effective response perceived by our hearing system. It plots the value of the response for each frequency during attack transients, instead of steady-state values. The block diagram of the proposed AQT method is reported below.

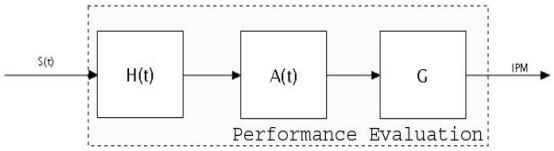

Figure 5: AQT Block Diagram

Where:
S(t) : AQT Stimulus signal.
H(t) : environment acoustic response to AQT Signal.
A(t) : extraction of quality parameters : Articulation, dynamic harmonic response.
G : estimation of the weight of these parameters on hearing.

Compared with original version of Liberatore, where the AQT stimuli is recorded, a great innovation, in terms of spare time during analysis, was obtained by introducing Virtual AQT [6]. This method calculates the AQT response performing a convolution between impulse response of the system, measured as shown in paragraph 1.1, and the AQT stimuli signal without recording it. An automatic tool was developed in order to quickly obtain AQT parameters, and to use them to synthesise a nice equalization filter shape. Moreover an automatic software tool was realized, which allows to perform AQT measurement with a user friendly GUI, and allows the user to synthesize a nice equalizer, fixing few degrees of freedom [9]. This software was named AQTTool. Starting from the eight measures of the impulse responses in the four fixed position (see 2.1), eight AQT curves were calculated for each car.

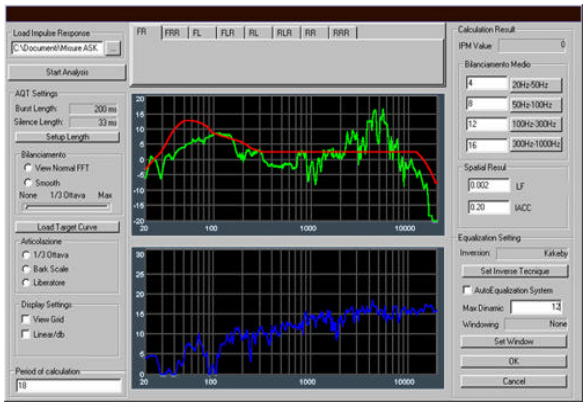

Figure 6: AQT Tool

These curves were averaged to obtain a single reliable and repeatable, highly-characterizing curve. Starting from this, a series of objective parameters were calculated.

4.2. Spectral balancing

A great number of those parameters was calculated as difference between a target curve, corresponding to most pleasant spectral response, and the AQT response of the car (as shown over). The target curve was chosen among a large set, evaluating the statistical correlation between subjective perception of spectral pleasure and their shapes. The preferred curve for automotive audio application resulted quite different from the standard flat curve, having a psychophysical trend, and resulted "coloured" during listening. The adherence of the response to this curve gives the best quality during assessment. Moreover different averaging algorithms were tested, exploring weighting and normalization procedures and psycho-acoustic post-processing inference.

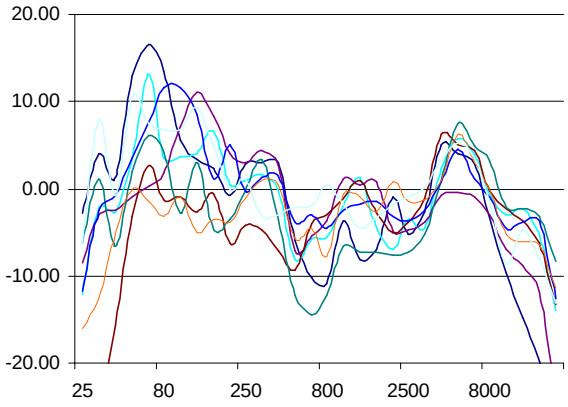

Figure 7: AQT responses of some cars

For instance a greater weight for low frequency was tested, for simulating logarithmic resolution of hearing system. The best one was fixed on the basis of correlation with subjective results.

4.3. Uniformity

The adherence of the response curve to target one isn't sufficiently descriptive of spectral perception. In fact, two cars with the same balancing should result very different in terms of pleasantness, if the differences from the target curve are differently distributed in the hearing range. It is preferable to have these differences equally distributed, in order to perceive a more balanced curve. An accumulation of the differences from target curve in a narrow bandwidth produces masking effects and affects the hearing.

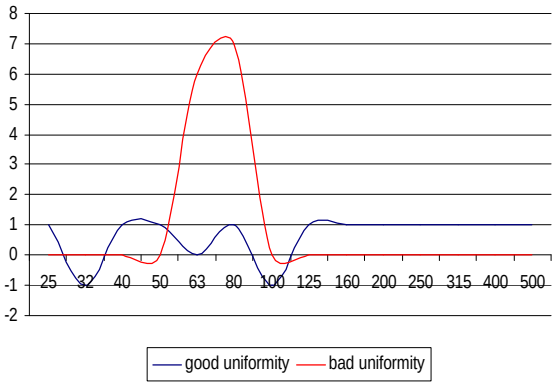

Figure 8: Uniformity

The formula (1) calculates the uniformity starting from the average deviation from target curve in three regions of the spectrum (2.)

$$\mathbf{U.}=\frac{S_{TOT}*\left(\left|S_{low}-S_{mid}\right|+\left|S_{low}-S_{High}\right|+\left|S_{mid}-S_{high}\right|\right)}{10} \quad (1)$$

$$S_{low}=\frac{\sum_{i=1}^{10}S_i}{S_{TOT}},\quad S_{mid}=\frac{\sum_{i=11}^{20}S_i}{S_{TOT}},\quad S_{high}=\frac{\sum_{i=21}^{30}S_i}{S_{TOT}},\quad S_{TOT}=\sum_{i=1}^{30}S_i,$$
$$se\ S_{TOT}\neq 0 \quad (2)$$

$$S_{low}=S_{mid}=S_{high}=0,$$
$$se\ S_{TOT}=0$$

4.4. Resonance indices

Calculated as a ratio between the max magnitude reached in low frequency (20-200Hz) and the energy in this region, this series of parameters gives an estimation of the tail inside the cockpit, due to resonances or uncontrolled behaviour of the loudspeakers. A typical effect called "boom effect" can be detected in this way.

4.5. Spatial Parameters

Two aspects of the spatiality can be highlighted:

- localisation of the sound sources;
- ambience and diffusion of the soundfield;

Both of them are traditionally estimated in theatres referring to the following parameters:

LE,LF,LFC which are spatial parameters calculated from response in B-format, recorded with Soundfield microphone. In this research these parameters were neglected.

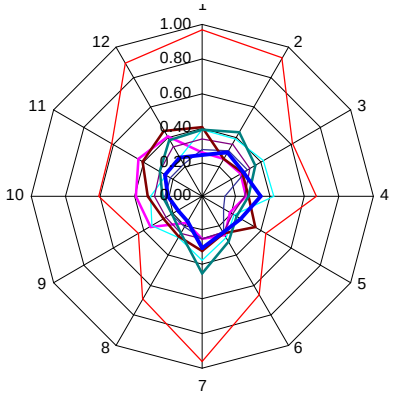

Figure: 9: IACC plot calculated from rotating signal

IACC: This parameter, introduced by Ando for the evaluation of spatial perception, is in use also in automotive field. The degree of stereophony was estimated by calculating the cross-correlation between the signals at the two ears. As already done in spectral characterization, the values obtained in the four measurement positions were averaged, and a stable and repeatable IACC was delivered. The results correlated well with subjective perception of this aspect of the sound field.

$$IACC = \max_{t} \left(\frac{\int_{t1}^{t2} p_{left}(t) \cdot p_{rigth}(t+t)dt}{\int_{t1}^{t2} p_{left}^{2}(t) \cdot p_{rigth}^{2}(t)dt} \right) \quad (3)$$

with -1ms < τ < 1ms

ITD: Acronym for Inter-aural Time Delay. This parameter can efficiently estimates the direction of provenience of sound. In this activity was adapted to automotive needs and a novel parameter, based on it, was developed. Particularly, a rotating signal elaborated in an anechoic room, was used as stimuli inside the ten car, and was recorded. The correlation between the original signal and the recorded ones resulted a valid indicator of the correct localisation of the sound sources inside the cockpit. So a high correlation means a good capacity of maintain the virtual source. Instead a low correlation indicates a “chaos” and phase instability inside the car. The novel approach to the analysis of the perceived audio quality inside the car is also able to estimate the acoustic performance of innovative surround systems, thanks to this parameter. In fact, using 5.1 or 7.1 system, a better localization of sound sources can be achieved, mainly from behind. So the index of performance presented here retrieves higher scores for new advanced system. In further research work, the validity of this parameter will be investigated with high accuracy.

4.6. Distortion Parameters

THD: classical parameters of distortion were inserted in the model during calculation of correlation. In particular THD, THD+N, and IMD were tested and the first one resulted the most correlating with subjective models.

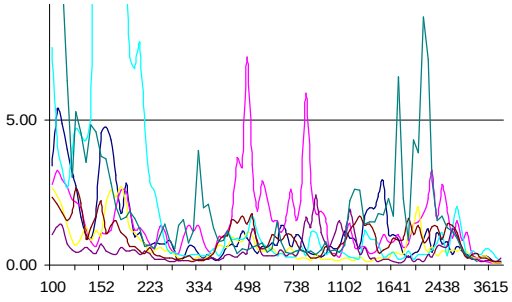

Figure 10: THD Responses

Only the odd harmonics are perceived as “distorting”, while even ones sometimes give an appreciated colour to the perceived sound. As it was done for spectral balancing, several weighting and normalization algorithms were tested, mainly giving importance to lower frequency where distortion is highly perceived.

4.7. Articulation parameters

This parameters are calculated directly by AQT Tool, and have been already explained.

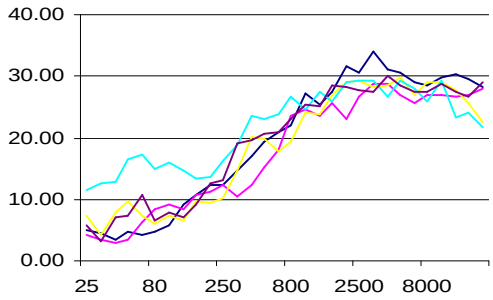

Figure 11: Articulation

4.8. Voice presence parameters

These parameters are calculated as energetic ratios between energy in speech region and overall energy of impulse response, representing the percentage of voice. A schematic voice filter was used.

At the end of the analysis, mainly aiming to optimisation of the correlation levels, only 6 parameters entered the model (plus one describing the "voice quality"). The more correlating parameters resulted to be the ones recorded at 80dB. So this level has to be considered the most pleasant for listening in the car without background noise. Also the distortion parameter recorder at 80dB resulted correlated with subjective models. So the dependence of it from recording level was investigated and resulted that a measure at 80dB is sufficiently representative of the behaviour of the system at high levels. In spite of the fact that at low level of reproduction the distortion is not perceived (typically under 5%), an excellent behaviour at low level corresponds to excellent behaviour at high level, and vice-versa. Indeed a strong correlation exists between the values of distortion at different SPL.

5. QUALITY INDEX

Summarising, the construction of the regressive "quality index" was conceived in three steps:

a) analysis of the correlation between "overall" audio quality judgment and single-feature verbal descriptors, that allowed for the deployment of the synthetic "like – dislike" evaluation into relevant components (that can be interpreted as the base psychological criteria which are used to mentally produce such an evaluation process), that is: construction of a "subjective model" of audio quality perception

quality evaluation = const. + $\Sigma_{i=1,\dots 4}\, a_i$*perceptual ftr $_i$

b) analysis of the correlation between single verbal descriptors and physical variables, for each quality feature in the subjective model (a linear combination of physical variables was constructed that showed maximum correlation with subjective response vs. minimum number of parameters)

perceptual feature $_i$ = k_i + physical parameter$_{i,1}$ +… + phys. par.$_{i,n}$

c) analysis of the correlation between the regressive "physical" representation of the verbal descriptors and the "overall" subjective quality judgment to be modeled (final measurable quality index, summarized as "IQSB": Indice Qualità Stereo di Bordo, that is "on Board Stereo Quality Index").

IQSB quality ind.= const. + $\Sigma_{i=1,\dots 4} a_i$ *[k_i + physical parameter$_{i,1}$+… + phys. par.$_{i,n}$]

6. RESULTS

In the following the results of the statistical analysis will be shown.
The results from the "naïve" and the "expert" panels have been treated separately and compared for validation and refinement of the model. More in detail, in step (a) it was found that, for "naïve" listeners, three descriptors could be retained (four, adding the "voice quality"), capable to efficiently explain the global judgment ("fidelity", "sound character", "bass/treble balance").
In pictures 12 and 13 the scatter plots of the linear combination of single descriptors vs. global evaluation,

both for 3 and 4 dimensional models, together with the correlation levels (for the naïve panel only).

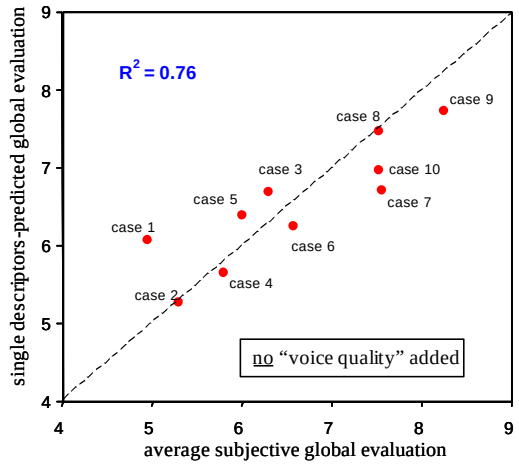

Figure 12: scatter and correlation of the subjective model relating 3 verbal descriptors to global evaluations, for the naïve jury panel (global evaluation = const. + a "fidelity" + b "sound character" + c "bass/treble balance") - *R^2 = 0.76, a < 10-3, RMSE = 0.55*

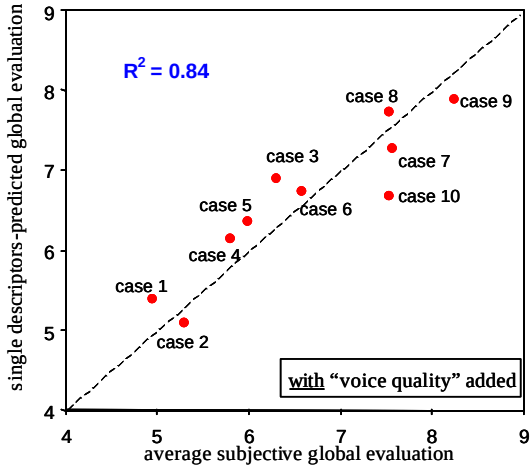

Figure 13: scatter and correlation of the subjective model relating 4 verbal descriptors to global evaluations, for the naïve jury panel (global evaluation = const. + a "fidelity" + b "sound character" + c "bass/treble balance" + d "voice quality") *R^2 = 0.84, a < 2*10-4, RMSE = 0.43*

The good correlation levels mean that three descriptors represent the main "perceptual dimensions" of the overall judgment, and that the use of the fourth further improves the model, so that the evaluation process as a whole will be efficiently emulated by a set of measures correlating with the 3 or 4 found dimensions.

b) In this step correlation was maximised between each main perceptual dimension and a linear combination of physical measures. It was found that ("Li" being the linear operator):

>> fidelity = L1 (corr_ITD_{an_sys} ; gap_TC)

>> sound character = L2(corr_ITD_{an_sys}; IACC; THD80)

>> bass/treble balance = L3 (res.; gap_TC ; unif.)

>> voice quality = L4 (sp_reg)

where:

- *corr_ITD_{an_sys}* is the correlation between ITD's (Inter-aural Time Delays) measured in anechoic conditions and inside the car-stereo system;
- *gap_TC* is the difference between a target dynamic acoustical response curve (identified in previous works) and the actual acoustical response curve measured inside the car compartment (corrected on psychoacoustic basis, taking into account effective masking patterns);
- *IACC* is a measure of the Inter-Aural Cross Correlation;
- *THD80* is a measure of the total harmonic distortion for an 80 dB test signal;
- *res.* is a "resound index";
- *unif.* is a uniformity index calculated for the actual dynamic acoustic response curve against the target curve;
- *sp_reg* is the spectral energy of the voice frequency region.

The correlation values between the average subjective scores of the individual verbal descriptors and those obtained by means of their measurable regressive models are the following: R^2 = 0.79 (a < 5*10-4 and

RMSE = 0.51) for "fidelity", R^2 = 0.9 (a < 10-4 and RMSE = 0.36) for "sound character", R^2 = 0.87 (a < 10-4 and RMSE = 0.48) for "bass/treble balance". As it has been said, the physical measure ("energy in the speech region") has been inserted independently from the corresponding subjective results, and its effect has been validated "a posteriori".

c) The calculated values of the regressive measurable models for each verbal descriptor were introduced in the subjective model (that is, with the coefficients obtained in step (a)) instead of the corresponding scores given by the (naïve) jury panel, and the correlation between "subjectively expressed" and "calculated" values was estimated, in order to confirm the robustness of the modelling procedure. As a matter of fact, the correlation obtained between jury scores and model calculated values (both for its 3 and 4 dimensional versions) resulted to be quite satisfactory, as it can be seen in pictures 14 (3 dimensional model) and 15 (4 dimensional model).

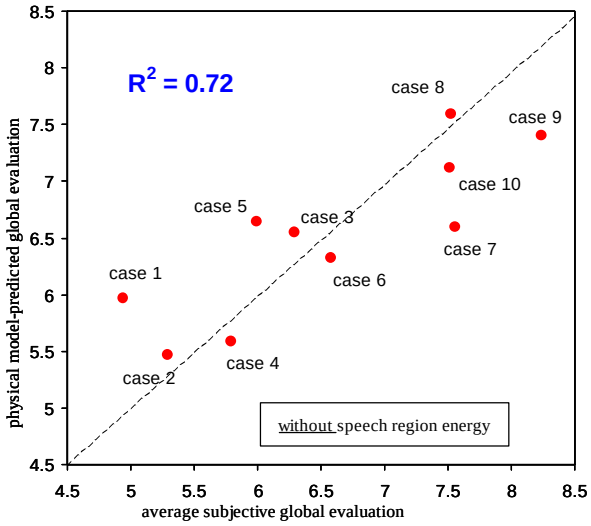

Figure 14: scatter and correlation of the predictive model relating 3 measurable partial indices to global evaluations, for the naïve jury panel
*R^2 = 0.72, a < 9*10-4, RMSE = 0.85*

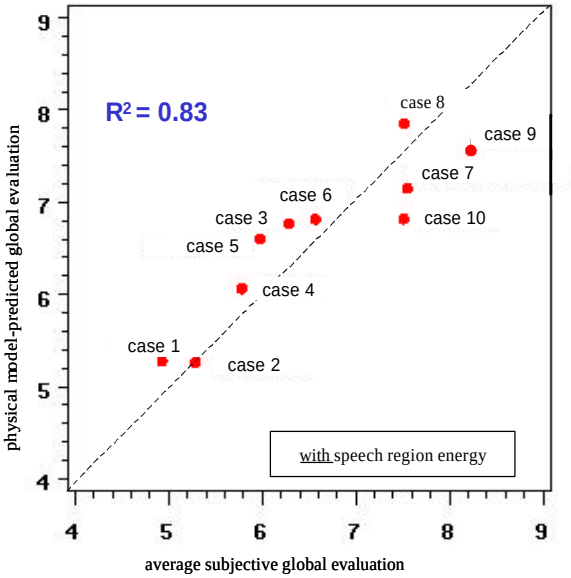

Figure 15: scatter and correlation of the predictive model relating 4 measurable partial indices to global evaluations, for the naïve jury panel
*R^2 = 0.83, a < 3*10-4, RMSE = 0.46*

The good correlation obtained with 4 variables suggests that the addition of the speech region energy improves the performance of the predictive model.

Subsequently, the results obtained with the 9 "experts" were used for the validation and refinement.

First the coherence between the performance of the two groups has been measured, again by calculating the correlation between their global evaluations. An R^2 = 0.85 (a < 10-4), ensures the average accordance between naïves and experts on the subject of "aural satisfaction".

Furthermore, the subjective model (developed in step (a) with the data of the naïve panel) has been applied, with and without the descriptor "voice quality", to the results of the listening test carried out with the experts. The accordance between the constructed model and the experts' judgment of global quality was measured by an R^2 = 0.91 (a < 10-4), for the 3 dimensional model, and by an R^2 = 0.95 (a < 10-4), for the 4 dimensional one, that is correlations even higher than those obtained with

the population used for finding the coefficients of the regressive model.

On the other hand, correlations between individual verbal descriptors scores for experts and naïves (R^2 = 0.48 and a = 0.03 for "fidelity", R^2 = 0.77 and a < 9*10-4 for "sound character", R^2 = 0.6 and a = 9*10-3 for "bass/treble balance") are quite low, except for the case of "sound character". This could mean that experts use the same categories as non-experts in evaluating sound, but they give them different meanings, probably more close to their underlying physical content (it seems to be coherent with the relatively high correlation corresponding with the locution "sound character", that doesn't relate to any particular acoustic "objective", directly measurable feature).

The same happens when correlating the physical models of the verbal descriptors (constructed with the data obtained by the naïve panel) with the scores assigned to them by the experts during the listening test: the physical definition of "fidelity" correlates with the experts' subjective scores with an R^2 = 0.42 and a = 0.04 , the one for "sound character" with an R^2 = 0.79 and a < 6*10-4, the one for "bass/treble balance" with an R^2 = 0.6 and a < 9*10-3 .

In any case, the correlation between the values of the predictive index (without and with the variable "speech region energy) and the overall quality evaluations of the expert panel showed an R^2 = 0.72 and a = 0.002 (with 3 dimensions) and an R^2 = 0.71 and a = 0.002 (with 4 dimensions).

This seems to point towards the hypothesis that the index is anyway suitable for predicting with sufficient reliability both "general public" and "expert listeners'" response, even though experts are probably more "skilled", more "precise" in giving stable evaluations, in particular for features directly linked to measurable properties they have some knowledge about.

Due to these considerations, the coefficients of the model (the expansion of the global evaluation in a linear combination of significant perceptual components, as in step (a)) have been re-calculated integrating the 9 experts panel together with the original 30 non-experts panel. The "refined" model proved not to be in contradiction with the results of the previous work, carried out with the 30 "naïve listeners panel", because the new coefficients lay within the statistical variation intervals calculated for the original ones (less than ±20% of the coefficient values). As a consequence, the correlation between predicted (i.e. without "speech region" parameters) and experimental scores increased from an R^2 = 0.76 (as given in picture X) to an R^2 = 0.9 (a < 10-4 and RMSE = 0.54), showing that the inclusion of a 25% "qualified" population (listeners that were assigned higher statistical weight) originated an improvement in the predictive performance of the index.

7. CONCLUSIONS

The IQSB, the measurable index developed in this work, proved to be a reliable tool for predicting at the experimental level the customer's feeling of quality for a given coupling of stereo components and the environment of the car compartment. It shows to be able to integrate the specialized performance of trained acoustic experts with the tastes of the general public. It opens the way to possible future refinements based on improvements in the inherent psycho-acoustic algorithms and it can be considered as a starting point for further developments in the direction of simulation-based "virtual audio design" methodologies. At present, the investigation of the perceived quality of a system with background noise has been performed, too, and results of subjective test session are under statistical analysis. In any case, IQSB can be considered, as today, as an effective solution for giving quick and objective answers to many car-stereo related target setting problems in which the automotive manufacturer must take decisions about the cost/performance ratio of new products, where up till now they were largely left to subjective, empirical evaluation.

8. FUTURE RESEARCHES

An accurate investigation of the effects of background noise on the quality assessment has already performed, with a session of subjective test. The results will be processed and presented in the future. Simultaneously the research on objective parameters and virtual listening room will continue.

9. ACKNOWLEDGEMENTS

The authors want to express their gratitude to ASK Industries (RE, Italy) for the fundamental support given

with the access to their special listening facilities and to the results of previous work on stereo reproduction quality. They also want to thank all ASK Personnel for their availability, experience and cooperation, in particular Dr. Ing. Carlo Brasca, who has always been an admirable host.

10. REFERENCES

[1] A. Farina, R. Glasgal, E. Armelloni, A. Torger - "Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music" - *19th AES Conference on Surround Sound, Techniques, Technology and Perception* - Schloss Elmau, Germany, 21-24 June 2001.

[2] M. A. Poletti - "A Unified Theory of Horizontal Holographic Sound Systems", *JAES Vol. 48, Number 12 p. 1049* (2000).

[3] A.Farina, E.Ugolotti – "Automatic Measurement System for Car Audio Application" – *104th AES Convention, Amsterdam,* 16-18 May1998

[4] A. Farina – "Simultaneous measurement of impulse response and distortion with a swept-sine technique",*110th AES Convention, Paris 18-22 February 2000.*

[5] I. Adami, F. Liberatore, "La messa a punto del sistema Diffusori-Ambiente", *Acustica Applicata srl, Via Roma 79, Gallicano – Lucca – Italy.*

[6] A. Farina, G.Cibelli , A. Bellini, "AQT – A New Objective Measurement Of The Acoustical Quality Of Sound Reproduction In Small Compartments", *AES 110th Convention Paper, Amsterdam 2001.*

[7] E. Ugolotti, G. Gobbi, A. Farina, " IPA – A subjective Assessment Method of Sound Quality of Car Sound System*" AES 110th Convention Paper, Amsterdam 2001.*

[8] A. Farina and F. Righini, "Software implementation of an MLS analyzer, with tools for convolution, auralization and inverse filtering", *AES 103 th Convention Paper, New York 1997.*

[9] A.Azzali, A.Bellini, E.Carpanoni, M.Romagnoli, and A.Farina, "AQTtool an automatic tool for design and synthesis of psychoacoustic Equalizers", *AES 114th Convention Paper, Amsterdam 2003*