Psychoacoustic Evaluation of Music Reproduction in Passenger Cars

H. Fastl[1], M. Ammler[1,2], A. Sulowski[2]
[1] AG Technische Akustik, MMK, Technische Universität München, Germany
Email: fastl@mmk.ei.tum.de
[2] Audi AG, Ingolstadt, Germany

Introduction

These days, almost all passenger cars are equipped with car audio systems of different sophistication. While physical measurements of characteristics like frequency response, sound pressure level, distortions etc. can be used as helpful engineering tools, the ultimate judgement about the quality of a car's audio system is performed by the customer with his or her hearing system. Therefore, it is feasible to check the quality of music reproduction in cars not only by physical measurements but in particular also by psychoacoustic studies.

Along these lines, a pilot study was performed assessing music reproduction in passenger cars. In essence, the subject sits in a standing car with the motor off and listens to a CD presented via the car's audio system. Several aspects of the performance are rated on a 10 step scale using a PDA. Results are discussed with respect to the sophistication of the audio system, possible gender differences, and biases due to product relationship.

Experiments

Up to thirty subjects aged between 24 and 48 years (median 29 years) participated in the experiments. Based on their self assessment, they could be regarded as "normal hearing" since none of them reported having hearing difficulties. 21 of them had some relation to the car company Audi, which was not the case for the remaining nine subjects.

For the psychoacoustic experiments, a CD was prepared with excerpts from music of different style like classic music, jazz, pop, rock etc. The excerpts had a duration of approximately one minute and were smoothly faded in and out. Excerpts were separated by pauses of five seconds duration. Based on instrumental as well as psychoacoustical evaluations, care was taken that the loudness of the different pieces of music was almost the same.

The music excerpts were selected to assess different regions of the audible spectrum, denoted Bass, Mid, and Treble. The Bass is subdivided into three frequency ranges, the Mid into four, and the Treble again into three ranges. As displayed in Figure 1, the whole audible frequency range from 16 Hz to 20 kHz is covered. However, more meaningful for psychoacoustic evaluations is a division along the Bark scale. Accordingly, the Bass extends from 0 to 2 Bark, the Mid from 2 to 16 Bark and the Treble from 16 to 24 Bark. For comparison the musical names of the (fundamental) tones of sounds from a violin are also given. The data displayed in Figure 1 reveal that the Mid range was chosen in such a way that it encompasses the fundamental frequencies of violin sounds.

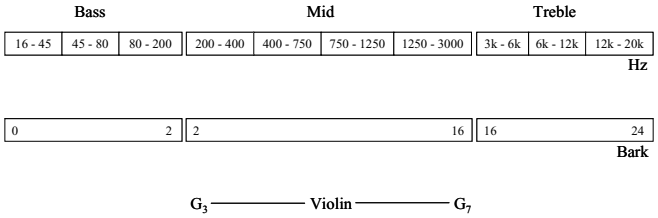


Figure 1: Division of the audible range in Bass, Mid, and Treble along a frequency scale in Hz versus a Bark scale. For comparison the musical names of the (fundamental) tones of sounds from a violin are also given.

For the experiment, the subject sat in a standing car with the motor off and listened to a CD presented via the car's audio system. Several aspects of the performance have to be rated on a 10 step scale using a PDA. Figure 2 shows some screenshots of the visual presentations on the PDA.

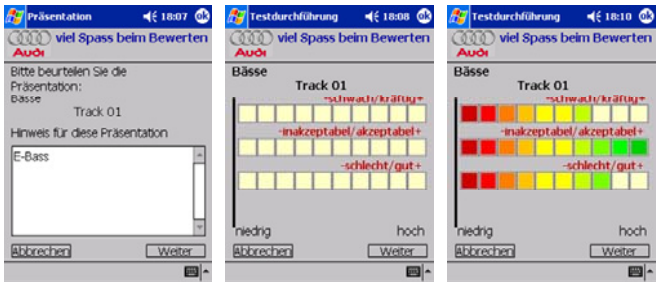


Figure 2: Screenshots from the PDA used for the evaluation of different aspects of the music presented.

The leftmost screenshot informs the subject that the presentation of track 1 on the CD concerns E-Bass. In the middle screenshot, three ten-step scales are presented denoted weak/strong, unacceptable/acceptable, and bad/good. The right screenshot shows as an example the rating by a subject: when the respective step is chosen, as a type of feedback its colour changes in such a way that positive ratings on the right end of the scale turn green and negative ratings on the left end turn red. Intermediate ratings are coded in yellow, and the whole system is somewhat reminiscent of a traffic light. For each track three to four aspects have to be evaluated. The responses are stored in the PDA and later transferred to a PC for data analysis.

Figure 3 enables an overview of the 32 aspects rated in total. While the first 31 ratings are related to different ranges on the Bark scale, item 32 is an overall rating of the quality of the sound reproduction by a specific audio system in a specific car without sound presentation.

testsequence	nr.	presentation	criteria	track	sound
1: sound evaluation	1	bass	-weak/strong+		
	2		-unacceptable/acceptable+	1	E-Bass
	3		-bad/good+		
	4	upper bass	-empty/full+		
	5		-booming/not booming+	2	String Bass
	6		-bad/good+		
	7	low mids (fundamentals)	-empty/full+		
	8		-dead/lively+	3	Pop 1
	9		-reverberant/clear+		
	10		-bad/good+		
	11	low mids	-empty/full+		
	12		-reverberant/clear+	4	Klassik 1
	13		-dirty/clean+		
	14		-bad/good+		
	15	mids	-colorless/colorful		
	16		-little fresh/very fresh+	5	Rock 1
	17		-bad/good+		
	18	upper mids	-dead/lively+		
	19		-indefinite/definite+	6	Rock 2
	20		-squeaky/not squeaky+		
	21		-bad/good+		
	22	low treble	-indefinite/definite+		
	23		-unnatural/natural+	7	Pop 3
	24		-little fresh/very fresh+		
	25		-bad/good+		
	26	treble	-dull/bright+		
	27		-obtrusive/unobtrusive+	8	Concert-Guitar
	28		-bad/good+		
	29	highest treble	-dead/lively+		
	30		-nerve-racking/not nerve-racking	9	Jazz 1
	31		-bad/good+		
	32	overall rating	-negative/positive+	-	without track

Figure 3: Overview of the 32 aspects assessed during the psychoacoustic experiments.

For the experiments, six different cars equipped with car audio systems of different sophistication were used. Table1 gives a ranking of the audio systems according to the technical effort (and price) involved. Three standard systems (01, 03, 06), a premium system (05) as well as two experimental concepts (02, 04) were used.

Vehicle	01	02	03	04	05	06
Audio-System installed	Std.	Concept	Std.	Concept	Premium	Std.

Table 1: Audio systems of vehicles 01 through 06

Results and Discussion

Figure 4 shows the overall rating for the quality of the audio systems installed in vehicles 01 through 06. Data are given as medians with quartiles derived from 30 data points each.

The data displayed in Figure 4 clearly reveal that the audio system installed in the concept vehicle 02 (white) gets the highest rating, and the standard audio system installed in vehicle 03 (yellow) the lowest rating. This is to some extent expected because of the significant differences in price. However, even at same price, the standard audio system installed in vehicle 01 (grey) is rated by two steps better than the standard audio system of vehicle 03 (yellow). The premium audio system of vehicle 05 (black) is rated somewhat better than the second concept audio system of vehicle 04 (red), although the price of the concept audio system is higher. Closer inspection reveals that for the rating of the second concept audio system in vehicle 04 (red) the inter-quartile range is larger than for the first concept audio system in vehicle 02 (white).

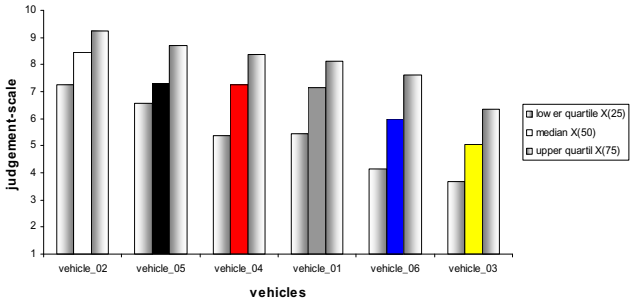


Figure 4: Overall rating for the quality of the audio systems installed in vehicles 01 through 06.

This means that the opinions about the quality of the concept audio system installed in vehicle 04 (red) are rather polarized. Similar effects show up for the rating of the standard audio system installed in vehicle 06 (blue), where even larger inter-quartiles are found.

Results displayed in Figure 5 give more detailed information about the perceptive features of the different car audio systems. The different aspects are scaled on the rays of a net, similar to a spider net, and for an optimum audio system, the goal is to encompass an area as large as possible.

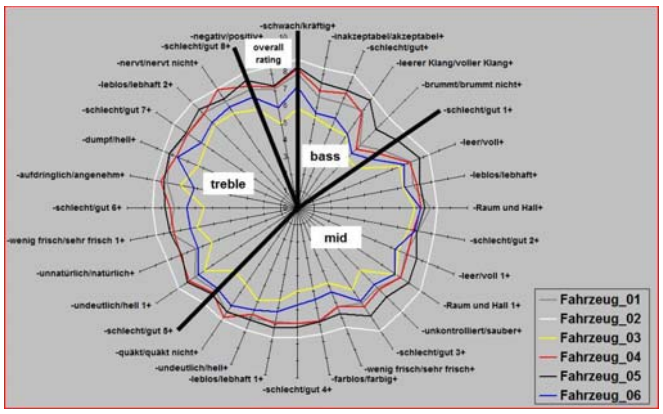
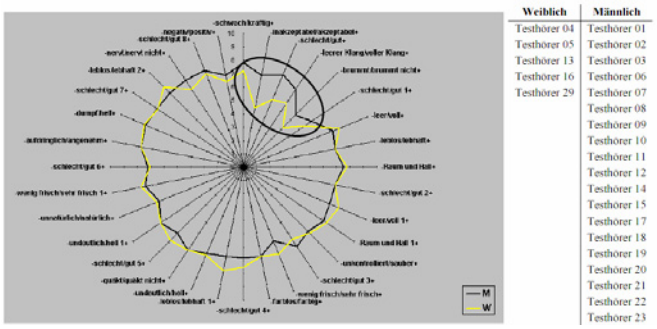


Figure 5: Rating of different aspects for the quality of the audio systems installed in vehicles 01 through 06, divided into bass, mid, treble, and overall.

The data displayed in Figure 5 reveal that vehicle 02 (white) encompasses the largest area of the net, and vehicle 03 (yellow) the smallest area. This result is in line with the overall rating as displayed in Figure 4. However, from the detailed rating shown in Figure 5 in addition it becomes clear that the net for vehicle 02 (white) is almost circular whereas the web for vehicle 03 (yellow) is rather rugged. This means that the first concept audio system installed in vehicle 02 gets high ratings in almost all features considered, whereas the standard audio system installed in vehicle 03 has several weak points. The premium audio system of vehicle 05 (black) shows also a more or less circular net, whereas the net of the second concept audio system installed in vehicle 04 (red) has significant drawbacks because of its booming *(brummt)*.

During the analysis of the psychoacoustic results, an interesting effect showed up when sorting the data according to the gender of the subjects. Corresponding data are displayed in Figure 6, where the yellow net represents results from female subjects and the black net results from male subjects.

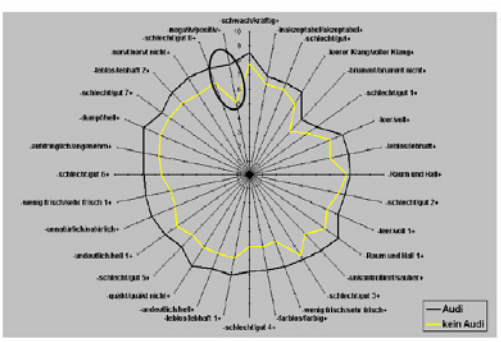


Weiblich	Männlich
Testhörer 04	Testhörer 01
Testhörer 05	Testhörer 02
Testhörer 13	Testhörer 03
Testhörer 16	Testhörer 06
Testhörer 29	Testhörer 07
	Testhörer 08
	Testhörer 09
	Testhörer 10
	Testhörer 11
	Testhörer 12
	Testhörer 14
	Testhörer 15
	Testhörer 17
	Testhörer 18
	Testhörer 19
	Testhörer 20
	Testhörer 21
	Testhörer 22
	Testhörer 23

Figure 6: Rating of different aspects for the quality of audio systems averaged across vehicles for female (yellow) versus male (black) subjects.

Although too few female subjects compared to male subjects participated in the experiments to draw statistically significant conclusions, the data displayed in Figure 6 indicate that the female subjects rated the reproduction of the bass more critical than the male subjects (see ellipse). In particular, the female subjects were annoyed by booming sound reproduction which rendered the whole bass region as unacceptable for them. On the other hand, in the mid and the treble region as well as in the overall rating, there is not much difference between the rating by female or male subjects.

Another interesting aspect showed up when dividing the subjects into two groups of people: the one group with members who have relations to Audi products because they work at Audi and/or drive an Audi car, and the other group with members who have no product relationship.



Audi	kein Audi
Testhörer 01	Testhörer 02
Testhörer 03	Testhörer 07
Testhörer 04	Testhörer 09
Testhörer 05	Testhörer 16
Testhörer 06	Testhörer 17
Testhörer 08	Testhörer 18
Testhörer 10	Testhörer 25
Testhörer 11	Testhörer 28
Testhörer 12	Testhörer 30
Testhörer 13	
Testhörer 14	
Testhörer 19	
Testhörer 20	
Testhörer 21	
Testhörer 22	
Testhörer 23	
Testhörer 24	
Testhörer 26	
Testhörer 27	
Testhörer 29	
Testhörer 31	

Figure 7: Rating of different aspects for the quality of audio systems averaged across vehicles for subjects who have some relation to Audi products (black) versus subjects with no product relationship (yellow).

From the total of 30 subjects, only nine have no product relationship, i.e. the group of subjects can not be divided into subgroups of equal size. Nevertheless, an interesting trend shows up: In almost all aspects displayed in the net, the subjects with no product relationship are about one step more critical than the subjects with product relationship. As concerns the overall rating (see ellipse in Figure 7), product relationship leads to a difference in rating of almost three steps (from ten)! Again, more subjects would be necessary to arrive at statistically significant differences. However, there is a clear trend that product relationship can lead to more "friendly" ratings. Therefore, as known from psychoacoustic experiments on sound quality, the choice of the group of subjects (experts, naive, biased etc.) can influence the results considerably.

Conclusion

In the present pilot study it could be shown that the proposed procedure for rating the quality of car audio systems, using a test CD for the music material and a PDA for collection of subjective evaluations, leads to reasonable results. After some final polishing of the procedure the next steps will be experiments in a standing car with the (idling) motor on. Further, similar experiments are necessary in cars running at different speeds on different types of roads. In this case, of course the driver can not use the original procedure, which, however, seems to be appropriate for passengers in the front seat as well as on the rear seats.

Acknowledgements

The authors wish to thank Dipl.-Ing. Florian Völk for editorial support.