Audio Engineering Society
Convention Paper 5595
Presented at the 112th Convention
2002 May 10–13 Munich, Germany

SELECTION AND TUNING OF HRTFS

Andreas Silzle[1]

[1] *AES Member, TC Electronic A/S, DK 8240 Risskov, Denmark,*
since February 2002 Institute of Communication Acoustics, Ruhr University Bochum, Germany, silzle@ika.ruhr-uni-bochum.de

ABSTRACT

The aim of the present work is the reproduction of a five-channel signal over headphones. Informal listening tests show that a large number of different HRTFs do not have the desired level of quality. The frontal localisation was either elevated or completely undefined. The coloration in all directions - even with correct IPTFs - was far too strong for a high quality reproduction. In order to overcome this problem two HRTFs sets together with IPTFs were selected out of a big database. These transfer functions were subsequently tuned by a tuning expert. The main methods used for tuning were smoothing and parametric equalizing of amplitude and phase with individual settings for every direction and for the left and right ear. Listening experiments that confirm the tuning results for a panel of listeners are presented and discussed. The resulting transfer functions have clearly reduced coloration and improved global localisation although with modest improvements in the frontal position.

1 INTRODUCTION

The aim of the development was the reproduction of a multi-channel signal over headphone. There should be an excellent out of head localisation together with no extra sound coloration[1] onto the original standard sound material. Figure 1

2 OVERVIEW

Here is a list of the paragraphs to get an overview: After the system definition in paragraph 3, follows the description of the problem of this development in paragraph 4. Paragraph 5 shows briefly previous HRTFs tests. Paragraph 6 describes the new method, that will be used here. The parameters of this investigation are listed in 7. The process of selection and tuning will be described in 8. The personal summary of the tuning results are presented in 9. An introduction to the listening experiments verifying these results are presented in 10. Paragraph 11 to 17 covers the listening experiments: coloration and localization tests without and with room simulation and quality tests with stereo and five channel sound material. After the final discussion in paragraph 18, the summary can be found in paragraph 19.

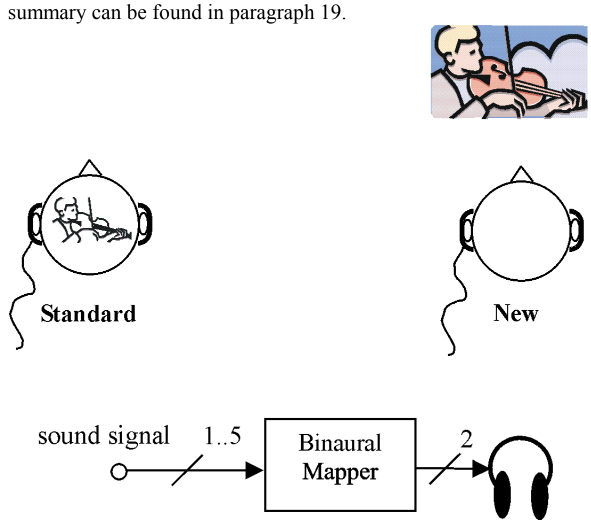

Figure 1: Introduction; Requirement: Out of head localisation with no extra coloration.

[1] Sound coloration is the change in timbre or spectrum with respect to a reference.

3 SYSTEM DEFINITION OF THE BINAURAL MAPPER

The standard system, which is here also the reference situation, is the unprocessed reproduction of sound material over headphone. The idea is the simulation of the five loudspeaker positions from the standardized 5.1 multi-channel set-up [1] over headphone. The new system should provide out of head localisation and no extra coloration. It should be realized with the following three processing blocks, Figure 2: First a good room simulation is necessary. Then the head related transfer functions (HRTF) make a reproduction over headphone possible, where the localisation is out of the head, compared to a normal headphone reproduction with a localisation inside the head. The inverse headphone transfer function (IPTF) has the task to make the "right" or "neutral" sound coloration.

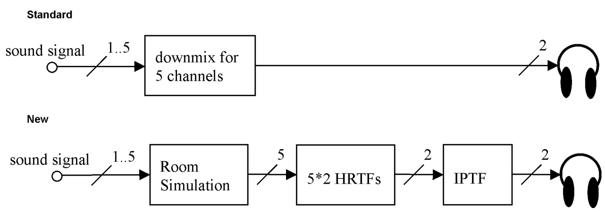

Figure 2: System Definition

4 PROBLEM DESCRIPTION

Normal binaural simulation is working with anechoic recordings. There the reference with respect due to localisation and coloration is the comparison with a reproduction in a real room. For the system defined here, the reference is the standard reproduction over headphones. The task was to change the localisation, getting it out of the head, but at the same time to keep the coloration or timbre of the sound, because it is ready mixed and produced. A sound reproduction over a good headphone has a similar sound coloration as a loudspeaker reproduction in a room. We introduce three more blocks into the transmission chain. They are needed to get an out of head localisation. Nor only do they provide the change in localisation, but they also change the sound coloration of the original signal. It is well known that every room and also every room simulation have their own sound coloration. It is also known that the magnitudes of the HRTFs and IPTFs are highly dependent on the frequency, and that they are both also highly individual between the different humans. For the planed application no individual HRTFs are possible and the "right" HRTF with the correct localisation should not have an extra coloration.
The following list is a result of informal listening of expert listeners to several dummy head recordings and simulations with different available HRTFs+IPTFs (collection see later).

4.1 Localisation

The simulation of an given virtual environment results in an imprecise localisation, at least clear elevation or even no out of head localisation with HRTFs (and reverb convolution) for the frontal direction. Dummy head recordings in real rooms results in the same results.
All directions can have a combination of several or all of the following negative points in the perceptual quality of sound localisation: different localisation directions between different persons, imprecise, smeared localisation, "connection" to the ears[2].

[2] This means, that there is beside the main localisation e.g. in the front additional localisation queus very close to the ears. Or with other words the localisation area is smeared out and reaches near to the ears.

4.2 Coloration

The HRTFs, even with the correct IPTF, produce a strong coloration compared to the original/standard signal.
The difference between measured HRTFs of different persons is big. It can be reduced by the introduction of the blocked ear canal measurement method [2]. But there are still differences of ±4dB in the frequency region around 5kHz and partly over ±8dB over 10kHz, Figure 3.
The difference between the headphone types and the uncertainty in measuring it (systematically and inter individual error) [3]) yields to a difference of above ±5dB at 5kHz between the IPTFs for different persons with the same headphone and ±15dB between different headphone types, see Figure 32. The author came with his own measurements to the same spread.

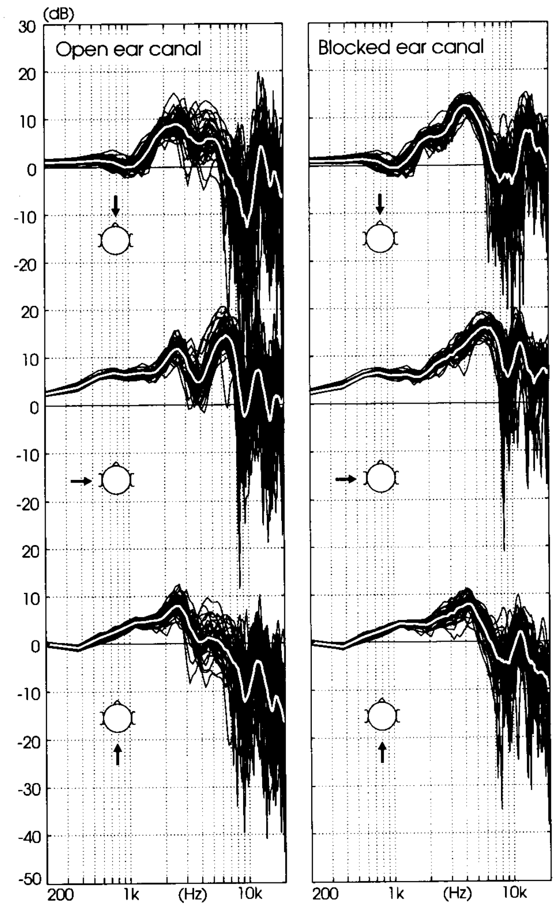

Figure 3 Comparison of left ear HRTFs for open ear canal (left column) and blocked ear canal (right colun). White curves represents means. Measured by AUC[3] [2]

Out of the author's experience also in an audio chain with HRTFs every single dB change of an equalizer in the high midrange will be audible and will change the sound coloration of the original signal, like in a normal mixing situation.
There is still the unproven but very interesting hypothesis, that the higher perceptual levels in the human perception use a match and inversion filter for localisation [4]. This makes the question more critical. Is there a non-individual HRTF+IPTF combination with

[3] Acoustic Department of Aalborg University, http://acoustics.auc.dk/

neutral coloration? Or is this match only achieved by individual HRTFs and all other approaches yield to an unacceptable coloration?
It seems, that localisation and coloration are not an uncoupled problem. What is now the right non-individual coloration setting for a binaural headphone reproduction?

5 PREVIOUS HRTFS TESTS

Møller et al. presented in [5] a “localization experiment, in which 20 subjects listened to binaural recordings from the ears of 30 humans.” In the critical median plane the error rate in real life was 16%, recording with a random subject 36%, with a "typical" subject 21%. Results were obtained with individual headphone equalization. It was a force choice test with a relatively rough resolution of 45^0 in the horizontal and vertical direction. An answer possibility for in-head localisation or localisation quality was missing. There was no question about coloration.
Møller et al. compared in [6], [7] and Minnaar et al. in [8], the localisation performance of different artificial-head and human head recordings with the real life condition. The procedure was the same as in [5]. The best artificial head (Valdemar from AUC) had an error rate of 37% in the median plane. Also the best human heads performed only with an error rate of 32% compared with the real life error rate of 10%.
Zhang et al. [9] used Kemar HRTFs “with the intent to refine them so as to exaggerate the perceptual differences for sound coming from different directions and to emphasize the pinna effects. With the refined HRTFs reduced front-back reversal rates and improved ability to locate moving sources are demonstrated.” They reported a degradation in coloration.
Spikofski et al. [10] showed the improvement in horizontal localisation by using the additional cue of head movement of a head-tracker. He showed also the vertical elevation of different dummy heads in the frontal region, in average 7^0, with head-tracker. And last but not least he showed the “Change of Sound Colour” for seven dummy heads and four different kinds of sound material. The broadband signal applause resulted for all dummy heads in a “clearly perceptible (2)” difference on five-grade impairment scale (5 = imperceptible).
RUB[4] measured also HRTFs with blocked ear canal, but 6mm inside compared to AUC. Hartung [11] showed the localisation performance for virtual sound sources and front-back error rates with individual HRTFs. The average error rate was 11.3%. There was no question about sound coloration.
Breebaart and Kohlrausch [17] “discuss the perceptual consequences of smoothing of anechoic HRTF phase and magnitude spectra.” “A first-order gammatone filterbank with bandwidths of 1 ERB is sufficient to describe the frequency dependence of both the phase and magnitude spectra.”

6 METHOD

The last two paragraphs indicate, that for the task given here and the aimed level of sound quality there is in the moment no measurement method and no selection of HRTFs and IPTFs, which gives a satisfying result.
The author will now introduce a novel process of finding and tuning a good HRTF+IPTF set. This process is perceptually guided: “what sounds right is right”. It is founded on the authors experience of years of tuning multi-channel audio coding (MPEG II) at IRT[5] and the experience of the involved tuning expert Ulrik Heise, who has tuned world class reverbs for many years at TC Electronic[6].

[4] Institute of Communication Acoustics of Ruhr University Bochum, www.ika.ruhr-uni-bochum.de/
[5] www.irt.de
[6] www.tcelectronic.com

There is the hope that there is a non-individual core in the HRTFs. The task was to find it.

The procedure was as following:
1. Real time selection out of a big number of HRTFs and IPTFs.
2. Tuning of the transfer functions with one tuning expert. Meanwhile controlling with few other expert listeners, to confirm different selections. This reduces the investigation time significantly compared to listening experiments and “sharpens” the results (no average taste).
3. The end results will be tested in a listening test.

7 PARAMETERS OF THE INVESTIGATION

The following tables show the collection of HRTFs and IPTFs for the investigation. They were collected both from humans and from dummy heads.

No.	Name	Az. res.	El. res.	Source
40	Human heads	22.5^0	22.5^0	AUC, not public
1	Dummy head Valdemar	22.5^0	22.5^0	AUC, not public
1	Dummy head Valdemar	2^0	2^0	AUC, not public
1	Dummy head Kemar asymmetric	5^0	10^0	MIT, [12]
1	Dummy head Kemar big symmetric	5^0	10^0	MIT, [12]
1	Dummy head Kemar small symmetric	5^0	10^0	MIT, [12]
1	Dummy head Kemar asymmetric, dc removed	5^0	10^0	MIT, adjusted
1	Dummy head Kemar big symmetric, dc removed	5^0	10^0	MIT, adjusted
1	Dummy head Kemar small symmetric, dc removed	5^0	10^0	MIT, adjusted
1	Sphere transfer functions	2^0	0	Calculated after [13]
6	Human heads	15^0	10^0	RUB, [14]
6	Human heads	17^0	11^0	TNO, [14]
61	Σ			

Table 1 HRTF databases used in the selection

No.	Name	Headphone	Source
19	Human heads	Beyerdynamic DT 990	AUC
41	Human heads	Beyerdynamic DT 770	AUC
41	Human heads	Sennheiser HD 560	AUC
41	Human heads	AKG K240	AUC
6	Human heads	Stax	RUB
6	Human heads	Sennheiser HD 520	TNO
1	Dummy head Kemar	AK K240	MIT
1	Dummy head Kemar	RSNova 38	MIT
1	Dummy head Kemar	Sennheiser HD 480	MIT

1	Dummy head Kemar	Sony Twin Turbo	MIT
1	Dummy head Valdemar	Beyerdynamic DT 990	AUC
1*3[7]	Dummy head Valdemar	Beyerdynamic DT 770	TC
1*3	Dummy head Valdemar	Koss HP/1	TC
1*3	Dummy head Valdemar	Philips in-ear plug	TC
1*3	Dummy head Valdemar	AKG 141 Monitor	TC
1*3	Dummy head Valdemar	AKG 240 DF	TC
1*3	Dummy head Valdemar	AKG K 66	TC
1*3	Dummy head Valdemar	Sennheiser 265	TC
3*3	Human heads	Beyerdynamic DT 990	TC
3*3	Human heads	Sennheiser HD 560	TC
3*3	Human heads	AKG K240	TC
>200	Σ		

Table 2 IPTFs used in the selection

The PTFs measured and reversed by the author are marked with TC. They were measured with the dpa microphone type 4060. The measurement method was the same as used by AUC with blocked ear canal. Most of the other measurements were done with the Sennheiser KE 4. The dpa microphone is of higher quality. An informal comparison showed, that the IPTFs measured with it, sounds "better" than the ones measured with the Sennheiser on the same head with the same headphone.

7.1 Sound examples

For tuning a broad band of different sound examples were used, which covers the whole frequency range and have quite different spatial impression to make the tuning process as independent as possible from the source material. Among them are:

- Mono and stereo signals from SQAM CD (several speech signals) [15],
- Several tracks from the Sound Check 2 CD [16]
- Five channel movie tracks from DVDs in DolbyDigital and DTS, e.g. Matrix, Titanic, James Bond.

7.2 Signal Processing

HRTFs and IPTFs were treated with a broad range of signal processing possibilities. Among them were:

- Parametric equalizing,
- Smoothing with different widths,
- Minimum phase calculation,
- Phase equalizing,
- Length limitation with different end windows.

The selection of the HRTFs and the signal processing was realized in Matlab, Figure 31. The processed transfer functions were downloaded to a DSP platform which did the FIR filtering with the audio signal. All calculations on the transfer functions and the audio processing were working in real time. The GUI included the visual feedback of showing the transfer functions (left and right) in time, magnitude, phase and group delay and separate the equalizer curves. The program included also the possibility to save and recall parts or whole sets of settings to give the possibility to switch immediately between them. This was important to detect also very fine differences between settings to find the "right way" during the tuning process.

[7] Number of repetitions

8 SELECTION AND TUNING

The difficult task for the tuning expert was:

1. Selection of a HRTF set, which was the best compromise with respect to localisation and coloration. The selection took part over the whole database. The search process was independent for every wanted direction.
2. Tuning this set to reduce the coloration (with respect to unprocessed) and sharpen the localisation, if possible.
3. Finding and tuning a general IPTF for a good overall performance for the different headphones.

9 SUMMARY OF THE TUNING RESULTS

Seen from the author and the tuning expert we achieve the following results:

1. A selection between the HRTFs is possible and necessary.
2. A individual tuning of these selected HRTFs for every direction and even for left and right is possible and results in a clear improvement – both for coloration and localisation. Even if the independent tuning of the left and right side of a HRTF pair is very fragile.
3. The really critical direction with respect to coloration changes is the frontal position. Here the more dominant cues of level and time difference between left and right are missing, so the whole perceptual information is coded in the coloration. The tuning expert can not find a setting for the frontal position, which reaches the same localisation quality as for all other positions on the horizontal plane. But he can find a much more neutral coloration as the different originals.
4. The human HRTFs sound in general better (more natural) than the dummy heads.
5. The IPTF has a main influence with respect to coloration. It is not so critical with respect to localisation. So there is no special IPTF necessary for every headphone type. The big difference in the transfer functions between headphones are more a matter of taste. They do not have a big individual influence on the different directions. The tuning can be done with a good neutral headphone and it fits also well to the others. The music studio standard AKG K240 Monitor was chosen. It sounds better like the nearly diffuse field equalized types like AKG K240 DF and Beyerdynamic DT990.

The details of the tuning process and the end result are not public. Parts of the work are under patent rights.

10 VERIFICATION OF THE RESULTS WITH SIX LISTENING EXPERIMENTS

A row of listening experiments test these selected and tuned transfer functions by a panel of listeners to verify, that also for this highly individual area of HRTFs a single tuning expert can tune for an average of people.
For all tests the reference was the unprocessed signal.
Most of the listeners have had some relationship to music, but they have had no experience in listening tests.

10.1 Parameters of all listening experiments

10.1.1 HRTFs

Source	Name
Best human head from AUC[8]	avh
Dummy head Valdemar from AUC	val
Kemar from MIT	kem
Improved AUC hrtfs	atc
Improved Kemar hrtfs	ktc

[8] Selected in a localisation test [5]

10.1.2 Sound signals

Relative dry mono signals and "normal wet" stereo and five channel signals.

10.1.3 Room simulation

The test was done with and without room simulation, to document the quality of the HRTF+IPTF combination alone and with room simulation. The quality of the room simulation algorithm and the settings had of course a big influence on the end result. And a pilot listening experiment indicated strong, that too much or the wrong reverb settings yielded to a strong negative voting of the listeners. But the whole field of room simulation or reverb setting was not part of this investigation.

10.2 Test software

The test design was inspired by the MUSHRA test [18]. It was a "double-blind multi-stimulus" test, with a given reference. The listener got a number of sequences with a number of sound examples per sequence. He could listen several times to every example, he could compare it with the reference sound and he could set the start and the stop point for the sound examples, so as to shrink it to the critical part only. He graded the examples with respect to the different questions. The listening test was also done with a matlab program, see Figure 4. There was an introduction page to explain the whole intention of the test and the grading scales. There was an introduction sequence to get familiar with the sounds and the GUI. Then the first sequence was not taken into account for the results, because the listeners took in average much longer for this sequence than for the other sequences (the time was measured), so it was a kind of trainings sequence. In every sequence was a duplicate track and sometimes whole sequences (in another order) were repeated to measure the confidence and consistency of the listeners. The test program collected all the votes and a second one calculated the result plots.

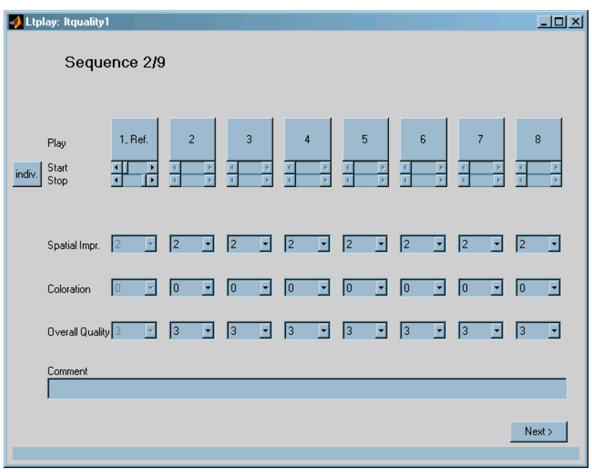

Figure 4 Listening test program, here with the example of the quality test 1.

11 COLORATION TEST 1

11.1 Parameters

HRTF+IPTF without room simulation.
Signal: mono female speech (SQAM track 49)

11.2 Question[9]

"Is the coloration very close to the reference signal grade (0). Is it better grade positive (up to +1), is it worse grade negative (down to -3)."

11.3 Result

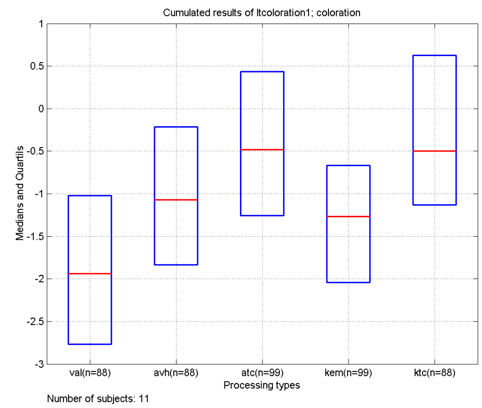

Figure 5 Over all directions cumulated coloration results for mono speech without room simulation

Significant differences between the medians are tested with the Kruskal-Wallis nonparametric one-way Analysis of Variance. The table shows the multiple comparison test result for a p-value of 0.05.

Tested	No significant difference	Significant difference
val	kem	avh, atc, ktc
avh	kem	val, atc, ktc
atc	ktc	val, avh, kem
kem	val, avh	atc, ktc
ktc	atc	val, avh, kem

11.4 Discussion

These are the cumulated results over all five tested directions (L, R, C, LS, RS). The number n in brackets behind the HRTF type at the x-scale is the number of answers for this boxplot. The middle line shows the median value and upper and lower line of the boxplot show the quartiles (50% of the answers are inside the quartiles). Median and quartile calculation and not average and standard deviation are used, because with only a five grade scale equidistant steps inside the scale cannot be assumed. Several test results also have clearly no normal distribution.

A comparison with a reference sound yields normally only a one-sided scale. During pilot experiments listeners report, they like the processed sound more than the reference, so a (+1) get introduced, to consider this internal reference.

The test shows the dummy head Valdemar is the worst due to coloration. Better are the dummy head Kemar and the human HRTF avh. The best are our two tuned HRTF sets atc and ktc. It makes no difference, if we start the tuning from Kemar or from AUC HRTFs. The tuned HRTFs came very close to the reference but also some listeners grade positive, the median value keeps at -0.5.

12 LOCALISATION TEST 1

HRTF+IPTF without room simulation.
Signal: mono female speech (SQAM track 49)

[9] Formulation as in the listening test introduction.

12.1 General Explanation

The following task was given to the subjects: “This is a localisation test. The localisation of the reference sound (at position 1 in every sequence) is in the head. Please grade the other examples. The localisation can only be very near the head (some centimetres), but the direction is interesting and also the perceived quality.”

12.2 Question 1: Direction

“Grading scale (5)..(1): Is the localisation at the intended direction (5) or far away from it (1). The intended direction is written at the top of the sequence.”
“Grade (5) for an error less than 10 degrees, (3) for an error of about 30 degrees, and (1) for more than 90 degrees. Grade (1) also for in-head localisation.”

12.3 Result

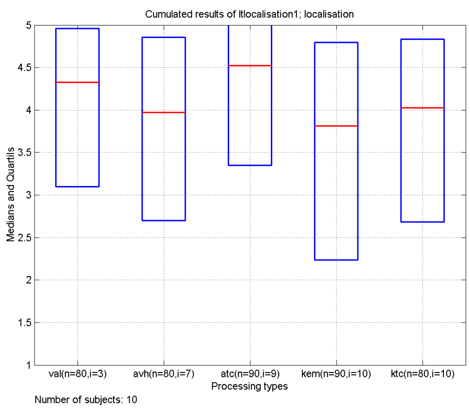

Figure 6 Over all directions cumulated direction results for mono speech without room simulation

No significant difference between all sets.

12.4 Question 2: Localisation quality

“Grading scale (5)..(1): Is the quality of the localisation very good (5) or very poor (1).”
“The grading here should be independent from the direction and the coloration. You should take into account quality parameters like blur and size of the localisation. One small sound source is the ideal. Grade (1) for in-head localisation.”

12.5 Result

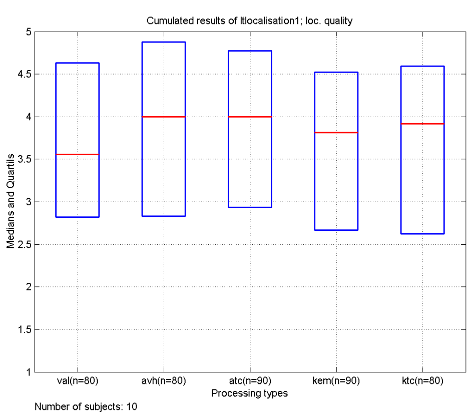

Figure 7 Over all directions cumulated localisation quality results for mono speech without room simulation

No significant difference between all sets.

12.6 Discussion

There was not the possibility to make a localisation test with external reference. For testing HRTFs without room simulation there is only an anechoic chamber as external reference possible. For the simulation with room simulation exactly this real room should be simulated. This was not intended. There was also not an external pointing device for the listener.
Because the investigation was only interested in the five loudspeaker positions of the 5.1 set-up (L,R,C,LS,RS), and only errors bigger than 10^0 or 15^0 were of interest, the method above described was chosen, where the listeners only get a paper which showed the set-up and they could decide with closed eyes about the position.
The number n in brackets behind the HRTF type at the x-scale is again the number of answers for this boxplot. i is the number of answers with the grade 1. This can be interpreted as the number of in-head localisations or completely wrong localisations.
Additional to the position question the author was interested in the localisation quality. Which is according to personal experience an important parameter.
The direction of the localisation shows no significant differences between the different HRTF sets. They are on a high level, most of the localisations are at the intended position or near by. Only for the 25% quartiles a difference can be seen. Here atc performs best. Related to the number of in-head localisations or completely wrong localisations Valdemar performs best. Important is, that the localisation for the tuned HRTFs are not worse than for the original, they are even slightly better.
For the localisation quality no significant difference can be seen. The result is on a high level but not on the best possible.

13 COLORATION TEST 2

HRTF+IPTF with room simulation.
Signal: mono speech (SQAM track 49)

13.1 Question

“Is the coloration very close to the reference signal grade (0). Is it better grade positive (up to +1), is it worse grade negative (down to -3).“
“Keep in mind, that you hear the reference without a room and the other signals in a room. Don't grade this difference. Also, don't take the sound source position in the room into account.”

13.2 Result

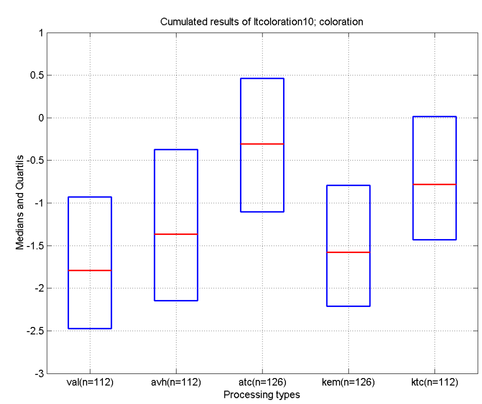

Figure 8 Over all directions cumulated coloration results for mono speech with room simulation

Tested	No significant difference	Significant difference
val	kem	avh, atc, ktc
avh	kem	val, atc, ktc
atc	ktc	val, avh, kem
kem	val, avh	atc, ktc
ktc	atc	val, avh, kem

These are the same significant differences as for the coloration test 1.

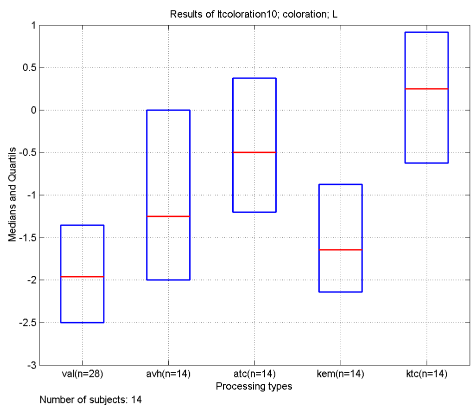

Figure 9 Coloration results for the left direction for mono speech with room simulation

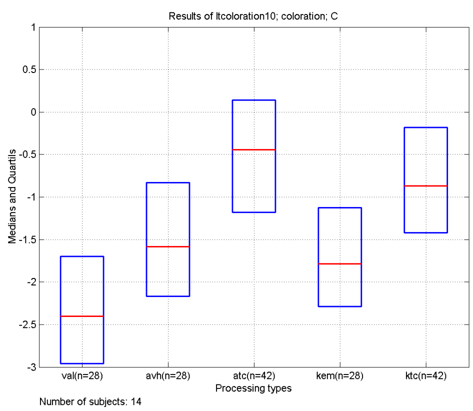

Figure 10 Coloration results for the center direction for mono speech with room simulation

13.3 Discussion

Figure 8..10 show the results with room simulation. The same room simulation setting is used for all HRTF sets.
From these figures can clearly be seen, that the ranking between the HRTF sets keeps with the room simulation like it was without room simulation. There is also no big difference between the single directions L, C or all cumulated directions.

14 LOCALISATION TEST 2

HRTF+IPTF with room simulation.
Signal: mono speech (SQAM track 49)

14.1 General Explanation

"This is a localisation test. The localisation of the reference sound (at position 1 in every sequence) is in the head. Please grade the other examples with respect to direction, distance and localisation quality."

14.2 Question 1: Direction

"Is the localisation at the intended direction (5) or far away from it (1). The intended direction is written at the top of the sequence. Grade (5) for a direction error less than 10 degrees, (3) for an error of about 30 degrees, and (1) for more than 90 degrees. Grade (1) also for in-head localisation."

14.3 Result

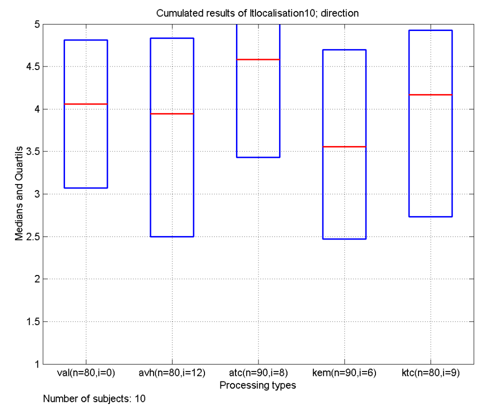

Figure 11 Over all directions cumulated direction results for mono speech with room simulation

Only atc and kem are significantly different.

14.4 Discussion

The results for the direction looks similar with room simulation and without room simulation. The median of atc is also here the best, but only significant compared to kem. Except for Valdemar all other HRTF sets have a number of completely wrong localisations (roughly around 10%).

14.5 Question 2: Distance

"Estimate the distance of the sound source in meters. (0) for in-head localisation. (4) for 4 or more meters distance."

14.6 Result

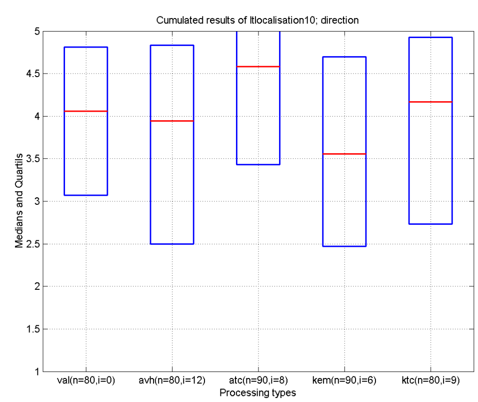

Figure 12 Over all directions cumulated distance results for mono speech with room simulation

Only avh is significantly different from kem.

14.7 Discussion

With the room simulation the question about the distance of the localisation can be added. The median for all HRTF sets is around 2.5m with a relatively small spread. This indicates that the distance is no problem and there seems to be no in-head localisation.

14.8 Question 3: Localisation quality

"The grading here should be independent from direction, distance and coloration. You should take into account quality parameters like blur, sound source width and how natural the room around the sound source seems. One small sound source is the ideal. Is the quality of the localisation very good (5) or very poor (1)? Grade (1) for in-head localisation."

14.9 Result

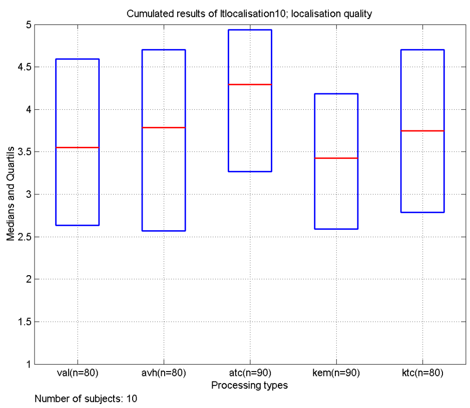

Figure 13 Over all directions cumulated localisation quality results for mono speech with room simulation

Only atc is significantly different from kem.

14.10 Discussion

The results for the localisation quality show a little bit more variance with room simulation than without. Atc's performance is better, but not significantly better.

15 CENTER CHANNEL

The following paragraph will give a closer look to the localisation results of the critical frontal or center position.

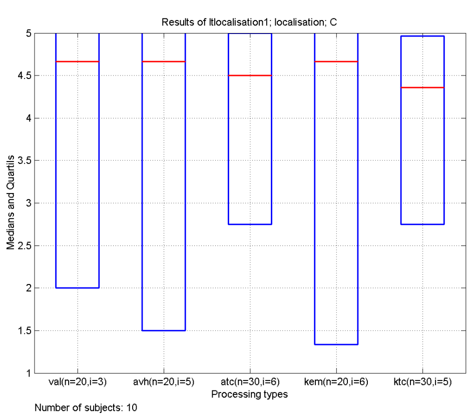

Figure 14 Direction results for the center direction for mono speech without room simulation

No significant difference between all sets.

15.1 Discussion

The median values are quite high, so most of the listeners perceive the center position at the right place. But the 25% quartiles are very low, the spread of a part of the answers is extremely high, about 20% of the grades are (1), which means, that in-head or completely wrong localisation occurs, and hence the so called front-back confusion can be assumed.
As mentioned before the tuning expert was not satisfied with his result for the center position and the test shows that the tuned HRTF sets result only in a small improvement compared to the original ones.

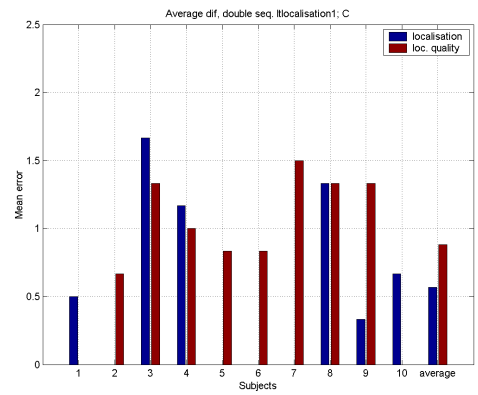

Figure 15 Average difference for a double sequence of the center for mono speech without room simulation

Figure 15 shows the average difference in the grading for the center position for every listener. The average over all listeners for the localisation question (left or blue bars) is 0.5, which means the listeners are relatively confident in their results. Listeners with no blue bar grade in both sequences exact the same. The average value is the same for the three with double sequences tested positions C,L,RS. This indicates, that the listeners are not more uncertain for the center position than for the others.
When we choose now only these "expert listeners", which have an error less than the average error, we get Figure 16.

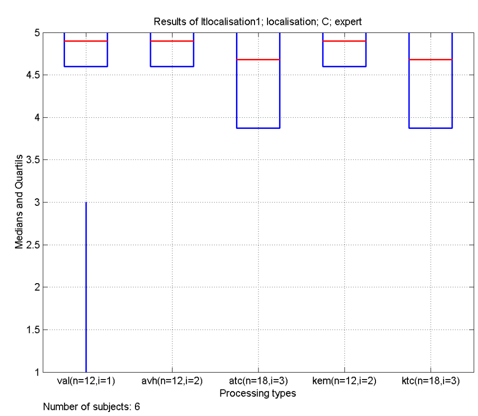

Figure 16 Direction results for the center direction for mono speech without room simulation for selected expert listeners

Figure 16 looks much better than Figure 14. There are only very few outliners (vertical line over val) and grades with (1).

The next figure shows the center direction with room simulation.

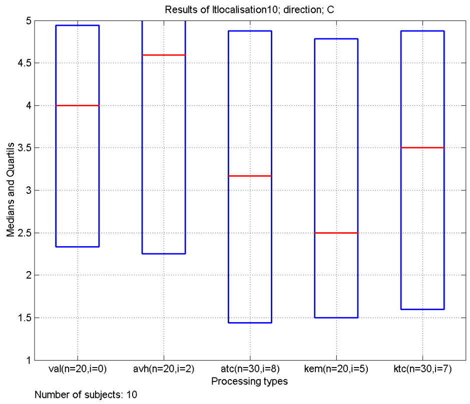

Figure 17 Direction results for the center direction for mono speech with room simulation

No significant difference between all sets.
The graph for the center position with room simulation looks even worth than the one without room simulation. The median values for atc, kem and ktc are quiet low. The author expected that the room simulation will help for the center position to track it out at the right position. The opposite is the case.
After making a selection with the same criteria as before for expert listeners we get the result of Figure 18.

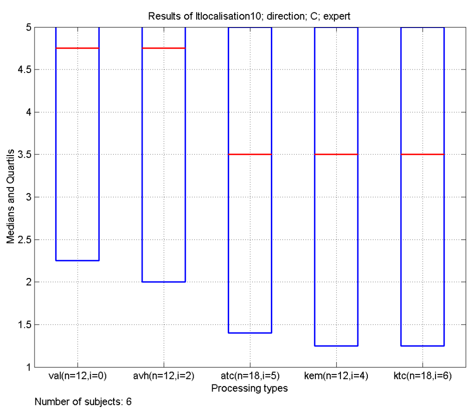

Figure 18 Direction results for the center direction for mono speech with room simulation for selected expert listeners

This selection don't "improve" the graph. Trying other ways for "defining" expert listeners don't make the result looks better. This shows clearly, that the localisation problem for the center position is not solved after the tuning.
There was no time in this work to investigate more into the influence of the room simulation on the center position.

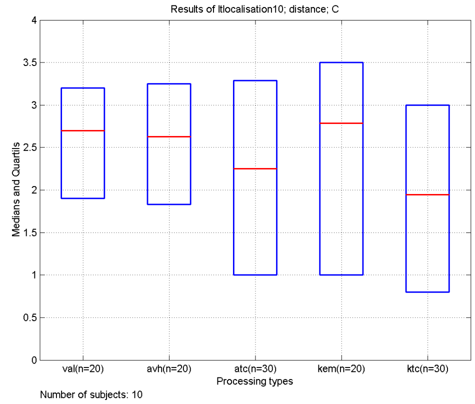

Figure 19 Distance results for the center direction for mono speech with room simulation

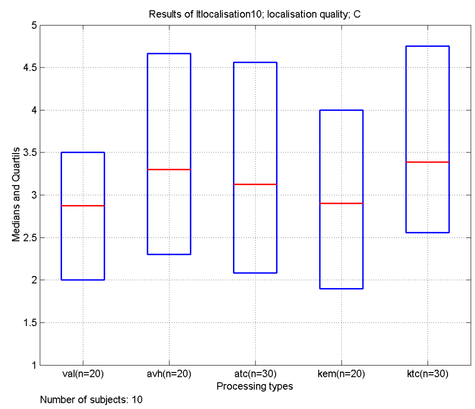

Figure 20 Localisation quality results for the center direction for mono speech with room simulation

The graphs for distance and localization quality give no further information for distinguish between the HRTF sets.

16 QUALITY TEST 1

The next test compare the reproduction of two channel "normal" music signals with room simulation.

Signal	Source	Genre
Verdi	SQAM, track 63, 0:22-0:31	Classic, Orchestra
Woodwind	Sound Check 2, track 81, 0-0:11	Classic, Ensemble
Limelight	Sound Check 2, track 83, 1:10-1:22	Rock

16.1 1. Question: Spatial impression

"Grading scale (1)..(5): Please grade your spatial impression.

5 very good	truly believable and engaging
4 good	like being in a small cinema
3 ok	better than normal headphone reproduction
2 no effect	as normal headphone reproduction
1 bad	worse than normal headphone reproduction

The reference is always the first track. It gets the score 2."

16.2 Result

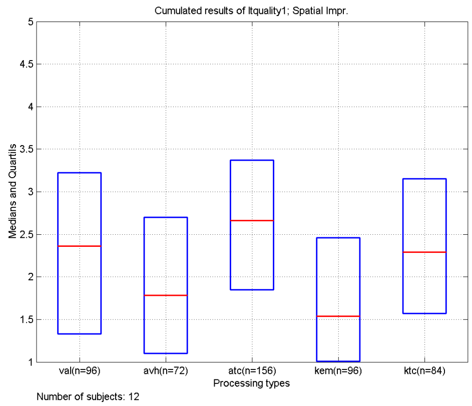

Figure 21 Spatial impression in the quality test for stereo signals

Tested	No significant difference	Significant difference
val	avh, atc, ktc	kem
avh	val, kem, ktc	atc
atc	ktc, val	avh, kem
kem	avh	val, atc, ktc
ktc	val, atc, avh	kem

16.3 2. Question: Coloration

"Grading scale (-3)..(+3): Is the coloration very close to the reference signal grade (0). Is it better grade positive (up to +1), is it worse grade negative (down to -3). The reference signal is the first sound of every sequence."

16.4 Result

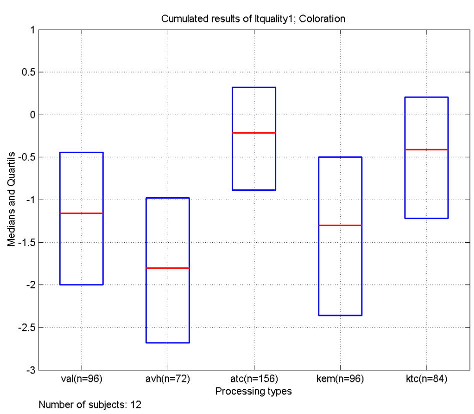

Figure 22 Coloration in the quality test for stereo signals

Tested	No significant difference	Significant difference
val	kem	ahv, atc, ktc
avh	kem	val, atc, ktc
atc	ktc	val, avh, kem
kem	val, avh	atc, ktc
ktc	atc	val, avh, kem

16.5 3. Question: Overall Quality

"Grading scale (1)..(5): Please grade the overall quality of the signals, with other words do you like it or not. For a very good quality grade with (5), for an average signal quality grade with (3), for a poor quality of the signal grade with (1)."

16.6 Result

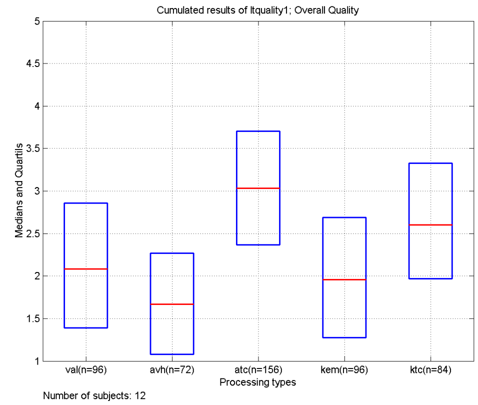

Figure 23 Overall quality for stereo signals

Tested	No significant difference	Significant difference
val	avh, kem	atc, ktc
avh	val, kem	atc, ktc
atc	ktc	val, avh, kem
kem	val, avh	atc, ktc
ktc	atc	val, avh, kem

16.7 Discussion

The answers for the stereo signals show for the question about the spatial impression no big difference between the different HRTF sets, Figure 21. The medians are around 2, so the listeners feel no big improvement compared with normal stereo and in-head localisation. For the coloration the differences are again bigger, Figure 22. Val is in this test better than avh. And the graph for the overall quality, Figure 23, looks very similar to the coloration graph. But also here the rates are not enthusiastic.

17 QUALITY TEST 2

The next test compares the reproduction of five channel "normal" music signals with room simulation.

Signal	Source	Genre
Aida	DVD	classic
Lovett	DVD	gospel
Matrix	DVD, AC3	movie, outdoor scene
Titanic	DVD, AC3	movie, indoor scene
Titanic	DVD, AC3	movie, outdoor scene

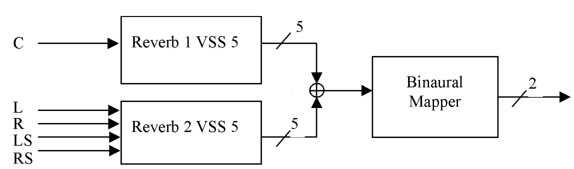

Figure 24 Room simulation Set-up

For the five channel test all five loudspeakers positions of the 5.1 set-up have to be simulated. This needs a room simulation with 5 separate source positions. The VSS 5 algorithm in the System 6000 from TC Electronic provides dedicated source positions for this. Every position has its own early reflection pattern. Because one algorithm calculates only four positions, two are used in parallel. The center channel is processed in the extra room simulation to get the possibility to make separate settings only for this position.

17.1 1. Question: Spatial impression

Same definition as for quality test 1.

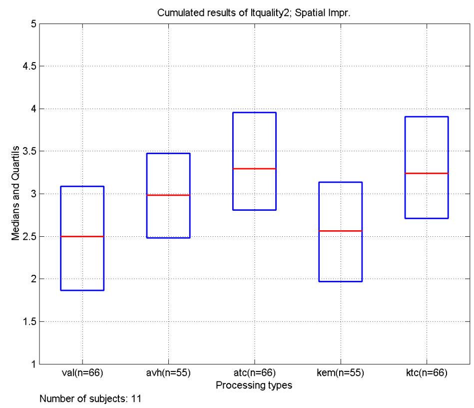

Figure 25 Spatial impression in the quality test for five channel signals

Tested	No significant difference	Significant difference
val	avh, kem	atc, ktc
avh	all	
atc	avh, ktc	val, kem
kem	val, avh	atc, ktc
ktc	avh, atc	val, kem

17.2 2. Question: Coloration

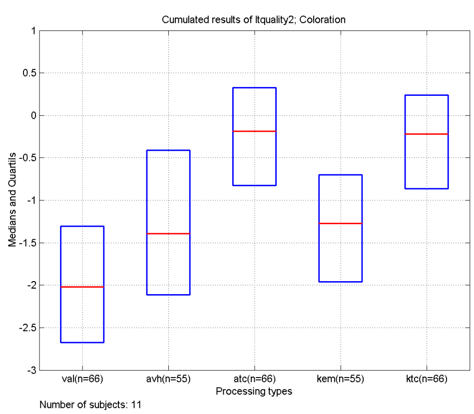

Figure 26 Coloration in the quality test for five channel signals

Tested	No significant difference	Significant difference
val	avh	atc, kem, ktc
avh	val, kem	atc, ktc
atc	ktc	val, avh, kem
kem	avh	val, atc, ktc
ktc	atc	val, avh, kem

17.3 3. Question: Overall Quality

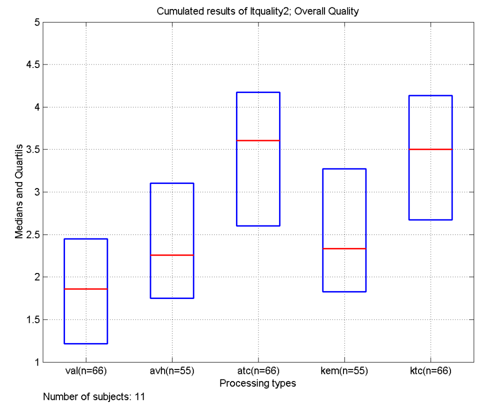

Figure 27 Over all quality for five channel signals

Tested	No significant difference	Significant difference
val	avh	atc, kem, ktc
avh	val, kem	atc, ktc
atc	ktc	val, avh, kem
kem	avh	val, atc, ktc
ktc	atc	val, avh, kem

Here the graph for all three questions: spatial impression, coloration and overall quality look very similar. atc and ktc are at the top, followed by kem and avh, and val at the end. In general the spatial impression rates and the ones for overall quality are higher for the five channel test than for the two channel test. The people feel a real improvement here. A five-channel signal holds much more spatial information than a two-channel signal, so it could be more convenient to have an out-of-head localisation.

18 FINAL DISCUSSION

Every sequence in the test has a duplicate track inside, to check the concentration of the subjects. There were also several duplicate sequences, to see the confidence of the subjects. According to this no listener gets “disqualified”. It shows also, that the number of steps for the scales are ok. The best listeners can grade with a confidence of about 0.5, some are a little bit over 1.

In general the listening test results confirm fully the selection and the tuning of the tuning expert. With respect to coloration the new HRTF sets are clearly better than the original ones and they are very close to the reference, the unprocessed signal.

With regards to the localisation, there was not much room for improvement, but even here the tuned ones are a little bit better than the originals. It is important that the coloration improvement doesn’t go in a line with a localisation degradation. The localisation answers have still a big spread. Beside of real different localisation, this can partly be related to the not optimal test method and also the inexperience of the listeners, to answer to localisation questions. For the critical center localisation only a small improvement can be shown in the listening test. But this was also not expected from the tuning expert. He just found no better solution. The answers for spatial impression show nearly the same improvements than the ones for coloration.

There was not the time to make the test with different headphones. There is only the personal experience form the author and the tuning expert, that the differences, that can be heard by the AKG 240, can also be heard with a wide range of other headphones, even with cheap in-ear plugs. So there is the hope, that no extra IPTFs for every headphone is necessary, but this should be confirmed in a listening test, especially for the critical frontal direction.

The settings and the level of the room simulation have an influence to the result. Also this needs further investigations.

18.1 Additional information

The loudness adjustment between the test items should be done very carefully. The next three figures of an additional test show clearly, that all parameters, which are tested here, are clearly level dependent. Only a relatively small change of ±3dB results in grading difference of about 0.5. Only two listener noticed, that it was the same example only at different level, and grade equal and make a remark.
Because the frequency distribution of the different HRTF sets are highly different, the loudness adjustment can only be done perceptual. It was done by the author and controlled by the tuning expert. We agreed inside 0.5dB. This was also for the author the level of just noticable difference.

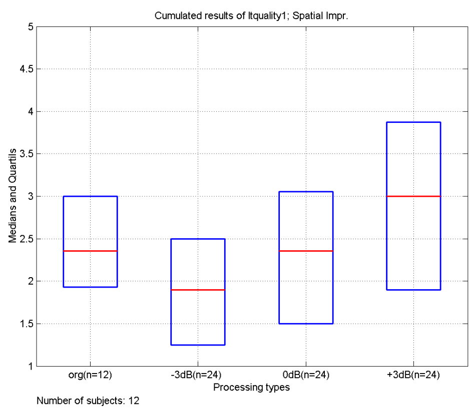

Figure 28 Spatial impression for the same signal with different levels

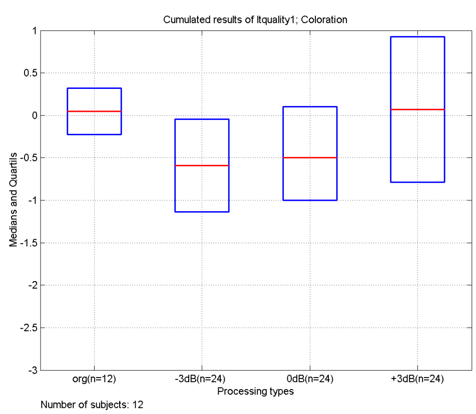

Figure 29 Coloration for the same signal with different levels

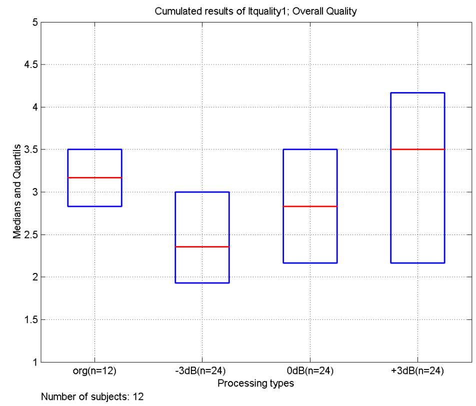

Figure 30 Over all quality for the same signal with different levels

19 SUMMARY

This investigation shows that there is a selection and tuning of HRTFs and IPTFs possible and necessary. The selection and tuning is done individual for every wanted direction. The tuning is possible even for the left and right side of a HRTF pair. The resulting transfer functions have clearly reduced coloration and improved global localisation although with modest improvements in the frontal position.
This process of selection and tuning can be done by one tuning expert and it gets confirmed in the listening test. So the perception in a system with HRTFs and headphone reproduction is less individual than expected.
The transfer functions are implemented in the Engage algorithm, implemented on the System 6000 from TC Electronic.

20 ACKNOWLEDGMENTS

I would like to thank the acoustic department from Aalborg University for providing us with a big number of HRTF and IPTF measurements and dpa microphones for lending us the in-ear microphones for the IPTF measurements.
I would like to thank Ulrik Heise for his excellent tuning job, his remarkable intuition, the fruitful collaboration and the energy he has put into this task, although this has produced some frictional loss.

21 REFERENCES

[1] Rec. ITU-R BS.775 Multichannel stereophonic sound systems with and without accompanying picture, 1994
[2] Møller, Henrik; et al.; Head-Related Transfer Functions of Human Subjects; JAES, Vol. 43, No. 5, 1995 May
[3] Møller, Henrik, et al.; Transfer Characteristics of Headphones Measured on Human Ears, JAES, Vol. 43, No. 4, 1995 April
[4] Theile, Günther; Zur Theorie der optimalen Wiedergabe von stereofonen Signalen über Lautsprecher und Kopfhörer (Theory about the optimal reproduction of stereophonic signals over loudspeakers and headphones); Rundfunktechnische Mitteilungen 25 (1981); Heft 4, S. 155-170
[5] Møller, Henrik; et al.; Using a Typical Human Subject for Binaural Recording; 100th AES Convention, 1996 May 11-14, Copenhagen, preprint 4157
[6] Møller, Henrik, et al.; Binaural Technique: Do We Need Individual Recordings?; JAES, Vol. 44, No. 6, 1995 June
[7] Møller, Henrik, et al.; Evaluation of Artificial Heads in Listening Tests; JAES, Vol. 47, No. 3, 1999 May

[8] Minnaar, Pauli, et al.; Localisation with Binaural Recordings from Artificial and Human Heads, JAES, Vol. 49, No. 5, 2001 May

[9] Zhang, Ming; Tan, Kah-Chye; Er, M. H.; Three-Dimensional Sound Synthesis Based on Head-Related Transfer Functions; JAES, Vol. 46, No. 10, 1998 October

[10] Spikofski, Gerhard; Fruhmann, Markus; Optimization of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics; AES 19th International Conference, Schloss Elmau, Germany; 2001 June

[11] Hartung , Klaus; Modellalgorithmen zum Richtungshören, basierend auf den Ergebnissen psychoakustischer und neurophysiologischer Experimente mit virtuellen Schallquellen; Dissertation, Bochum, 1997

[12] Bill Gardner and Keith Martin, HRTF Measurements of a KEMAR Dummy-Head Microphone, 1994; http://sound.media.mit.edu/KEMAR.html

[13] Duda, Richard O.; Range dependence of the response of a spherical head model; JASA 104(5), p. 3048-3058, November 1998

[14] HEAD acoustics, TNO, Ruhr Universität Bochum, SEXTANT Avionique; AUDIS catalogue of human HRTFs (CD), http://www.ika.ruhr-uni-bochum.de/; 1998

[15] EBU; SQAM CD, http://www.ebu.ch/tech_t3253.html, 1988

[16] InterStudio; SOUND CHECK 2 CD; http://www.interstudio.co.uk/sndchk.htm

[17] Jeroen Breebaart, Armin Kohlrausch; The perceptual (ir)relevance of HRTF magnitude and phase spectra; AES 110th Convention 2001 May 12-15 Amsterdam

[18] Stoll, Gerhard; Koyamernik; Franc; EBU listening tests on Internet audio codecs; EBU Technical Review; June 2000

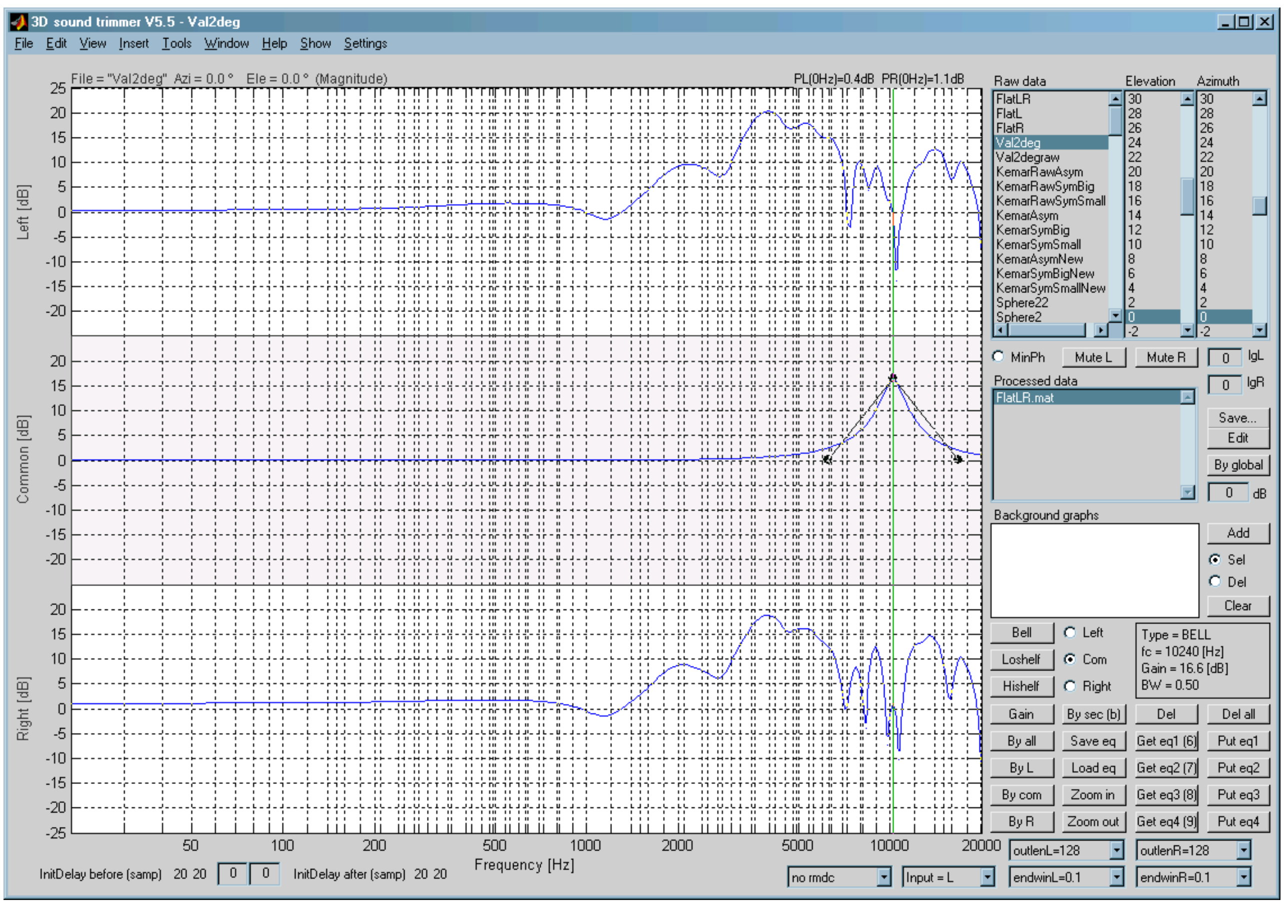

Figure 31 Matlab GUI of the selection and tuning program.

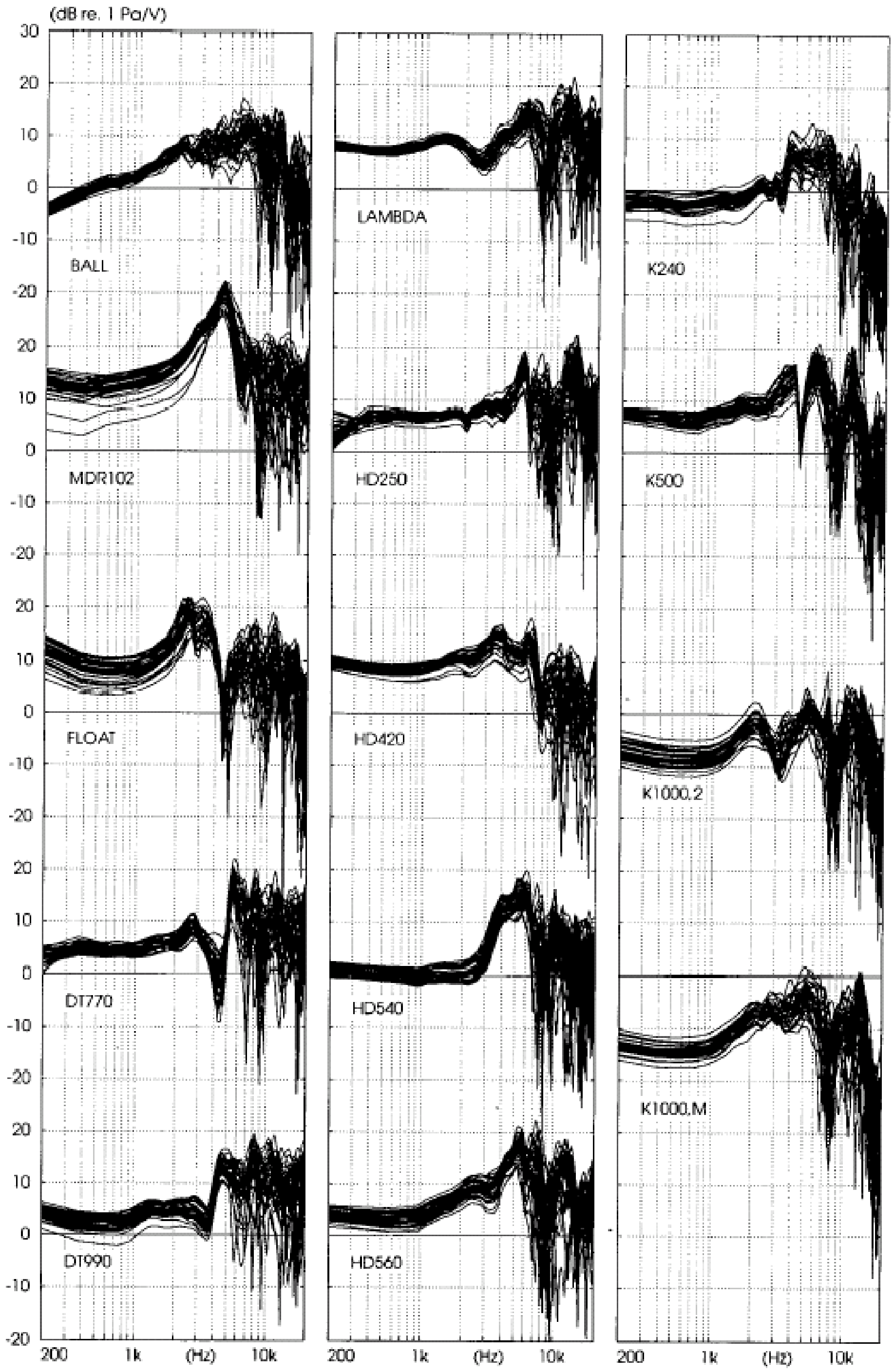

Fig. 6. PTFs for 14 headphones measured at the blocked ear canal of 40 human subjects.

Figure 32: Headphone transfer functions, measured by AUC [3]