ACÚSTICA 2000

TECNIACÚSTICA MADRID - 2000

"SWEET SPOT" INVESTIGATION OF CROSSTALK CANCELLATION

PACS REFERENCE: 43.60.Lq

Schmitz, Oliver [1]; Amate, Joan [2]
1 Institute of Technical Acoustics, Technical University of Aachen
Templergraben 55,
52056 Aachen. Gemany
Tel: 49 241 80 79 84
Fax: 49 241 88 88 214
E-Mail: osc@akustik.rwth-aachen.de; joan@bcntelecom.net

ABSTRACT

To reproduce binaural signals usually headphones are used. Although the reproduction of 3D-sound works quite well this way, some disadvantages can be found, which negatively affect the immersion. To overcome these disadvantages a system based on crosstalk cancellation is under development. In this paper results of listening tests are presented performed to answer the question "What is the maximum movement or head rotation away from the sweet spot, until a noticeable change is perceived?"

INTRODUCTION

Binaural signals are often used in virtual reality systems to generate a three dimensional soundscape. To reproduce these signals usually headphones are used. Although the reproduction of 3D-sound works quite well this way some disadvantages can be found. Mainly in-head localization and the discrepancy between the perception of a spatial sound event on one hand and the fact of wearing headphones on the other hand. Both problems negatively affect the immersion in the virtual environment. To overcome these disadvantages another technique is under development, which intends to replace the headphones with a pair of loudspeakers. This technique is based on crosstalk cancellation (CTC)[1,2]. Current investigations are focussed on a dynamic crosstalk cancellation system, which allows the listener to move freely. This requires to steer the point of maximum channel separation (also called "the sweet spot") according to the listener's position and orientation. In order to do this, different questions have to be answered. How can the crosstalk cancellation filters be synthesized? How often are new filters to be calculated? Or in other words: What is the maximum movement or head rotation away from the sweet spot, until a noticeable change is perceived? How to switch between filters? The first question has already been discussed by Gardner [3] in some extent. To answer the second question listening tests have been performed. This paper discusses the results of these tests.

LISTENING TESTS

First, a typical localization test was carried out, similar to many others carried out in papers about 3D sound spatialization. This test will later be used as a reference to evaluate the performance of a dynamic system. It will not be discussed in detail in this paper. The second test is intended to determine in which area the listener can move without detecting any change in perception of a single virtual source, that is, where the CTC filter works properly. The goal in this test is to know when a dynamic crosstalk cancellation system should change to new filter coefficients in order to keep the perception of virtual sources independent of the listener's movements.

Both the excitation signals and the crosstalk cancellation filters used in this test are generated based on measurements using our custom made dummy head. This dummy head is an asymmetric head model with different pinnae for the left and for the right ear. The excitation signals were generated by placing a loudspeaker at the source position and recording with the dummy head. These recordings took part in a hemi-anechoic chamber. The speaker used was a studio monitor. The stimuli for the tests each consisted of five white noise bursts of 250ms, each burst separated 500ms from the other. The speaker setup, as seen in Figure 1, consists of two speakers at ±45° with a distance to the reference point of 2,4 m. The tests are performed in free-field conditions, in the hemi-anechoic room. To obtain the crosstalk cancellation filters the dummy head was properly positioned in order to make its ear canal positions coincident with the ear canal positions of a mean test person in the test. From the measured transfer functions between the loudspeakers and the dummy head, four FIR filters of 413 coefficients each are calculated and stored in a DSP system to perform the crosstalk cancellation.

A set of 12 people took part. They were all unfamiliar with sound localization experiments and presented no audiological dysfunction. All of them were instructed to be honest in their judgements, instead of taking the experiment as a localization "skill test".

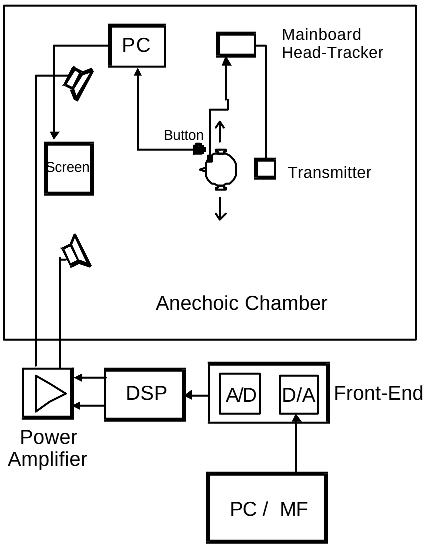


Figure 1: Setup for the working area test

Three different sequences were used for the working area test:

- Signal #1: Alternated white noise bursts of 300 ms with elevation 0° and azimuth +90°/-90°.
- Signal #2R: Alternated white noise bursts with elevation 0° and azimuth +45°/+90°.
- Signal #2L: Alternated white noise bursts with elevation 0° and azimuth -45°/-90°.

The reason for choosing this signals is the fact that a source at 45° azimuth and 0° elevation matches exactly with the real loudspeaker position. According to the "power panning" property of the CTC systems, when a virtual source is exactly located at the position of one of the loudspeakers, all the power is reproduced by this loudspeaker. The 90 degree direction on the other hand is known to be the most critical azimuth direction. So although the listener walks away from the sweet spot, he will always have a reference source at 45°, and can focus on the changes at the source placed at 90°.

In the working area test, the listener stands in front of the loudspeaker setup. He wears a head tracking sensor mounted on a pair of glasses. This allows to determine the subject's head position and rotation at any time. The data from the tracker is send to a computer which displays the position of the subject in a simple graphical interface. Front-back and left-right translations are displayed as well as rotations around all three axes. This helps the subject to monitor his head position relative to the reference position, and to recover the reference point or "sweet spot" easily. The first step in this experiment is to train the subject to be familiar with the head tracking system. Once familiar with this environment, the subject listens to Signal #1 while keeping the reference position. The reference point at the beginning of the test is determined by the dummy head position when recording the transfer functions. This position is obtained by simply putting the glasses on the dummy and recording the position with the tracker. At this point, the subject knows which kind of signal he is listening to and is asked whether he can localize it well or not. Depending on which of the two sources can be localized better test signal #2L or #2R is used for the further investigations. The listener is asked to move in different directions, focusing the attention on the source at 90°. The listener was instructed to seek for the first change in perception of the source at 90° when moving. As soon as noticed, he pressed a button several times. This makes the connected PC record the coordinates of the current head position relative to the reference point. Further the subject reports the kind of the perceived change. Then, the subject is asked to move further in the same direction until a second kind of change occures. This is recorded the same way as before. Note that the subjects are instructed to try to move in only one direction, keeping the others constant with the help of the graphics.

To describe the investigated movements a coordinate system is introduced. The x-axis points to the front, the y-axis points to the left, and the z-axis points upwards. In order to supply an easy nomenclature to refer to the three possible head rotations, the following convention will be used:

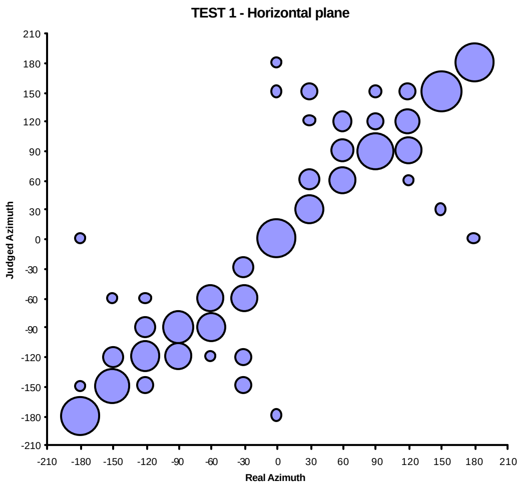
- Rotation over Y axis: “Yes” movement. It is the rotation used to say “Yes” with the head.
- Rotation over Z axis: “No” movement.
- Rotation over X axis: “Maybe” movement, in analogy with the two precedents. This rotation consists of bending the neck as if bringing the head over the shoulder.

The movements proposed to the subject in this test are:

- Translation: Left, right, front and back.
- Rotation: “No” movement and “Maybe” movement.

The translation along the z-axis (up and down) is not considered in this test. Although it has some influence, it is not a critical movement for the CTC system. When moving the head up and down, the interaural level differences remain the same, and so do the interaural time delays. As the experiment is done in free-field conditions, no other reflections but the own from the subject’s shoulders are to be considered, and of course these are invariant with the head translation along the Z axis. Moreover, depending on each subject’s height, different reference points will be recorded, as the relative height between the hearer’s ear channels and the loudspeakers is influencing the measurement. Concerning head rotations, the “Yes” movement is also not considered for this study. It is known that this rotation is influencing the crosstalk cancellation, as the position of high frequency notches in the HRTF spectrum is changing. However, it was decided to leave this rotation out of the study because it is difficult to isolate from other movements. In addition to this, the test is long enough and exhaustion of the test subjects should be prevented.

LOCALIZATION TEST RESULTS



The results of the localization test are only presented for completeness. Figure 2 shows histograms for the judged azimuths at each target azimuth on the horizontal plane for all subjects. The size of the circles is proportional to the number of events produced at each position. The ideal error-free plot would be maximum size circles placed along the y = x diagonal. Localization in the horizontal plane is quite good, and comparable with results of other authors. Typical front-back reversals are found as well as a typical response variation and response bias. Some problems occurred in the median plane, which are under further investigation.

Figure 2: Histogram for the localization of the source in the horizontal.

WORKING AREA TEST RESULTS

A summary of the results of this test is shown in Table 1. For all the movements tested, results for each of the directions were obtained. In order to present the results in a compact way, the absolute values were taken and the direction information is omitted (i.e. for lateral translation the movements to the left and to the right are processed together). So, only one value will be obtained for each movement.

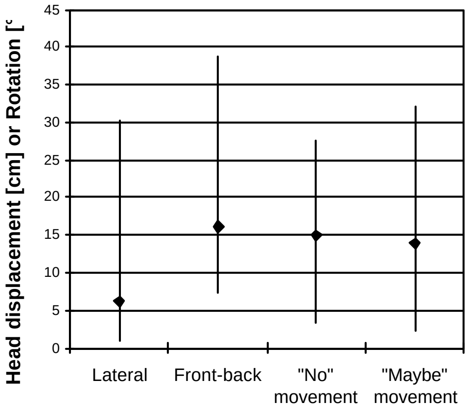
For the lateral movement, a total of five subjects stated that the source moved backwards (to an azimuth position higher than 90°). This is the most frequent observation. Other judgments are that the source is moving forward, a change in coloration occurres, and that the source becomes more diffuse. For some subjects these changes happened for quite small excursion from the reference position as shown in Table 1. A high sensitivity against this movement is expected because the interference responsible for the cancellation is strongly sensitive against changes in phase, which is of course depending on the relative time delay of the interfering signals. Nevertheless it is a bit surprising that the smallest lateral offset reported was less than 1 cm. This would make a dynamic crosstalk canceller almost impossible to implement, because the average accuracy of usual head tracking devices is around or even above this value.

When moving forwards and backwards, less sensitivity to changes is expected because of the constant relative time delay and the low sensitivity of source and head directivity against this movement. This is supported by the results of the test. The subjects can step much further away from the reference point than they can with lateral translations. The perceived changes observed here are even more diverse than for the lateral translation. Five of the subjects detect coloration changes but only three report position changes. The sensation experimented in general was that the source stays by the subject's side along the translation, which is the required behavior because of the fact, that binaural signals are head related. Three subjects could even move far away from the sweet point without reporting significant changes in the source characteristics.

Type of movement	Median	Min	Max	Std dev
Lateral	6,2 cm	0,83 cm	30,2 cm	8,61 cm
Front-back	16,14 cm	7,24 cm	38,73 cm	8,78 cm
„No" movement	15,04°	3,35°	27,55°	7,66°
„Maybe" movement	14°	2,26°	32,11°	9,39°

Table 1: Summary of the results obtained in the working area test

Going deeper into figures, as shown in Figure 3 for the lateral translations a median value of the data around 6 cm is obtained. The diamonds represent the median values and the lines the range of values. The median value is chosen instead of the mean because the set of values obtained is quite spread, although a certain coincidence is expected. Due to the spreading of the data, it is hard to get a statistical validation of the results. Some people are capable of reporting changes in the source by just moving slightly, whereas others aan step far from the sweet point without reporting any change.



The median value for the front-back translation is around 16 cm, which means a higher headroom for the dynamic crosstalk canceller. Below 7 cm none of the subjects detected a change of the source. This minimum value is almost 10 times larger than the one for lateral movement. For this reason the front-back movement will not be discussed any further. For the implementation of a dynamic crosstalk canceller the minimum allowed movement is critical. Thus, a closer look into the results is necessary. Figure 3 shows some histograms for lateral translation and the two rotations investigated. In this histograms the quantity of subjects is displayed over the distance or angle interval where they detect a change of the source.

Figure 3: Working area test results. Diamonds show the median value and lines represent the minimum and maximum value stored.

The two observations below 1 cm, which are a bit worrying for the idea of dynamic crosstalk cancellation, require some more examination. Aside of the head tracker problem discussed above, if changes were detected below 1 cm, the speed needed for filter calculation and switching in a dynamic implementation would be almost unreachable.The existence of groups 2 and 3 is clearly related to a different criteria when considering a change of the source. Naturally, there is a lot of subjectivity in this observation and it is hard to unify the criteria of the subjects. However, the observations may be influenced by the following conditions:

a) The test person can not really stay at a fixed position. A subject asked not to move still moves with an amplitude of approximately 2 cm.
b) There is some latency time between the detection of a change of the source and pressing the button.

In the case of the "No" movement, the most frequent event is the travelling of the source from 90° to 45° (reported by 8 of the subjects). Only two subjects reported that the source moves to the back of the listener. Only in one case the listener perceives no change in the source when rotating the head. The standard deviation for this experiment is lower than in the experiments working with translation, although it is still high. As in the previous experiments the data acquired is rather spread. In the histogram shown in Figure 4 it is difficult to determine a tendency, as the values are distributed along a big interval. Again, the need for criteria unification is stated. But fortunately the minimum rotation reported is above 3°.

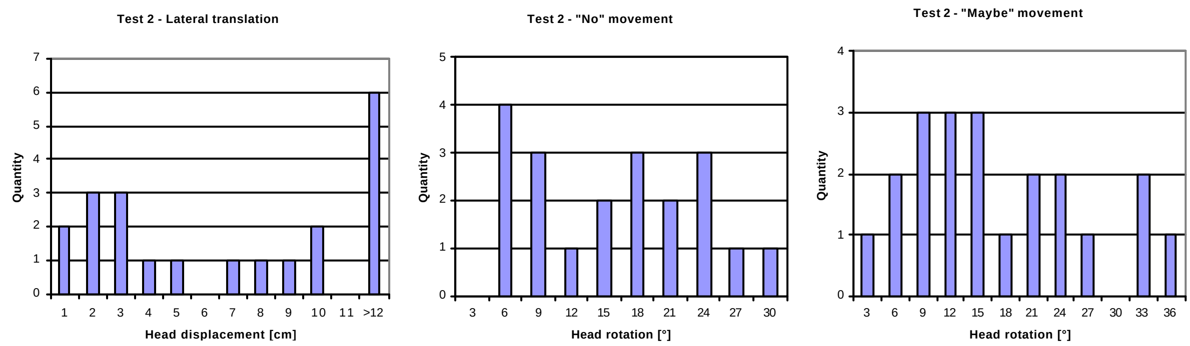


Figure 4: Histogram for lateral translation the “No” and the “Maybe” movement. Bars show the quantity of observations on each interval.

The “Maybe” movement also leads to a degradation of the crosstalk canceller performance. An important characteristic of the transmission path is the early reflection coming from the subject’s shoulders. With the movement proposed, the distance ear-shoulder is altered and so are the transmission paths. As the difference between the direct ray and the first reflection is about 0.5 ms, a comb filter is produced by the sum, creating a notch at certain frequencies. When the distance between the shoulder and the ear changes, so do the comb filter frequencies, leading to a mismatch between the crosstalk filter corrections and the actual corrections needed. So a lack of performance of the crosstalk canceller is expected together with this head rotation. A wide variety of events such as the translation of the source to the front and to the back (8 observations), and the diffusion of the source (3 observations) are reported by the test subjects. Other perceptions are that the source changes its coloration, elevates or moves downward and reduces the level. The fact that some subjects report that the source changes its height is similar to the situation observed when rotating the head over the horizontal plane. The listener’s impression changes depending on what he expects.

Again spread values of rotations are recorded. In this case a median value of 14° is obtained from the data of rotations on left and right sides. It is hard to get a statistical validation of the results. More subjects and more repetitions of each experiment would be needed to get more significant data. This would also reduce the figures of the standard deviation. As stated in House and Shively[4], working with trained people is a way to reduce variance and increase repeatability in this kind of test. In our case, untrained people were submitted to the tests. Although all of them first passed a short training before each test, all of them were unfamiliar with localization tests.

REPEATABILITY TEST

All the subjects were asked to search the exact point where they could perceive changes of the source. For this purpose, most of them needed to move repeatedly in the test direction, but only one point was recorded. So, no repeatability evaluation can be extracted from this experiment. To study the repeatability of the working area test another test was done. Using the same setup with only two trained subjects, familiar with 3D spatialization sound systems. Only two of the movements proposed in the working area test are studied now: translation along the Y-axis (moving left and right) and rotation over the Z-axis (“No” movement). The subjects are asked to look first for a coloration change and then for a localization change on the source. For each event, a minimum of five points are recorded, starting always from the reference position. Each time a point is recorded, the subject recoveres the reference position to eliminate the memory of the previous position.

The results of this experiment are summarized in Table 2. As for the translation movement both subjects can distinguish coloration from localization changes, it is not so for the rotation movement. While Subject 1 could only perceive coloration changes, Subject 2 distinguished better localization changes. So for the rotation, “1st event” means coloration change for Subject 1 and localization change for Subject 2.

From the data obtained it can be stated that the standard deviation for the points recorded by one subject is highly reduced compared against the figures of standard deviation across subjects A-L.

However it is hard, the subjects 1 and 2 could look repeatedly for the points where they noticed a change with certain accuracy. On the other hand, there are discrepancies between the mean values of the data sets recorded by Subject 1 and Subject 2. Comparing this data with the median values of the previous test there are also discrepancies. However, a tendency can be observed both in translations and rotations for subjects 1 and 2. About 2 cm translation and 10° rotations are the mean values observed, which are lower than the values of the previous test.

		Subject 1		Subject 2	
Observation		Mean	Std Dev	Mean	Std Dev
Moving left	Localization	-1,94 cm	1,11	-3,24 cm	1,26
	Coloration	-1,31cm	0,92	-2,04 cm	0,95
Moving right	Localization	2,61 cm	0,95	3,73 cm	0,58
	Coloration	2,48 cm	1,03	3,62 cm	1,38
Rotating left	1st change	-10,01°	3,24	-3,60°	2,13
Rotating right	1st change	8,23°	2,13	10,45°	1,06

Table 2: Results of the repeatability experiment

This experiment shows that the problem in the working area test, which has high values of standard deviations, is not the intra-subject deviation but the inter-subject. We define intra-subject deviation as the standard deviation of the experiments done on one subject only (several realizations), and inter-subject deviation as the standard deviation of the experiments done on various subjects (one realization per subject). When a test subject chooses a criteria to determine some change in the sound source, it is able to repeat quite accurately this observation. However, each subject has different hearing characteristics and criteria, which makes the inter-subject deviation high. Hearing characteristics are dependant on physical conditions: pinnae, ear-shoulder distance, ear's frequency response, etc. and the criteria depend on psychological factors. There is a lot of subjectivity involved in the perception of a undetermined change of a source, so each listener will have his own criteria of what a "change of the source" means.

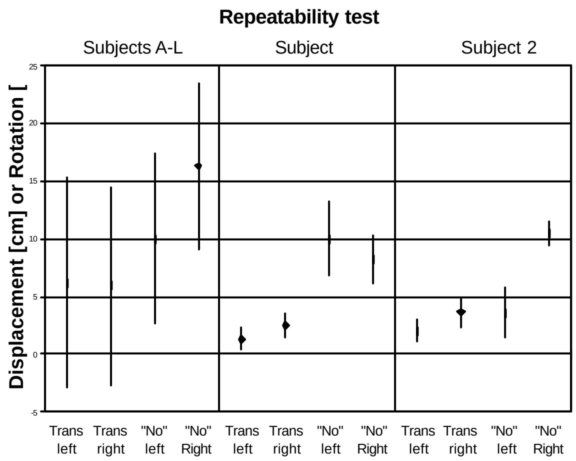


Figure 5: Repeatability discussion. Left shows the results of the previous test, with averaged data over 12 subjects, center and right show the results of the repeatability experiment.

CONCLUSIONS

In the working area test the problem of the high standard deviation in the subjects' observations arises. The set of values obtained should be treated in a conservative way, expecting always the optimum values to be under the median values there obtained. An approximation to the possible working values is obtained in the repeatability test, although this is only a starting point which needs to be verified. Unifying the observation criteria of the test subjects is needed in order to reduce the variance of the experiment, although this should not condition the listener observations. It is hard to unify the criteria without conditioning the test subject; if clues of what is expected to happen with the source are given to the listener, then he is psychologically ready to perceive this, even if this event is not really true. Through the last test it is observed that these subjective criteria are quite accurately kept by each subject. Nevertheless, still high differences between the judgements of each subject are observed. More specific tests should be carried out, in order to look for more exact figures. The results of the tests will now be used for the first implementations of the dynamic CTC approach. A final evaluation can anyway only be made with the dynamic system.

[1] Bauer, B.B.; Stereophonic Earphones and Binaural Loudspeakers; J. Audio Eng- Society, 9(2) 1961, pp. 148-151
[2] Schroeder, M.R. and B.S. Atal; Computer simulation of sound transmission in rooms, IEEE Int. Conv. Record, 7 1963, pp.150-151
[3] Gardner, W. G.; 3-D Audio Using Loudspeaker; Phd. Thesis MIT Media Lab 1992
[4] House, N. and Shively, R.; The Placebo Method, a comparison of in-situ subjective evaluation methods for vehicles, AES 108th Convention 2000, Paris; preprint 5136