

Marshallplan Scholarship Report

-

**INVESTIGATIONS ON TEMPORAL,
SIMULTANEOUS AND
SPECTRAL MASKING FOR
DEVELOPING A 3D AUDIO CODEC**

Christian Kranzler

August 31, 2011

Abstract

In this paper we present a study conducted within a Marshallplan Scholarship at the Department of Cognitive and Neural Systems (CNS), Boston University. Spatial release from masking under several circumstances was investigated. A study was carried out with 12 participants, 6 of them were tested using a high frequency masker and a low frequency target, 6 of them were tested using a low frequency masker and a high frequency target. It turns out, that with the low frequency target no masking occurs, but with the high frequency target masking is present and so is binaural release from masking.

Contents

1. Introduction	5
1.1. Spectral (simultaneous) masking	5
1.2. Temporal masking	5
1.3. Spatial masking	5
2. Methods	7
2.1. Setup	7
2.2. Participants	9
2.3. Procedure	9
3. Experiments	13
3.1. Experiment 1	13
3.2. Experiment 2	13
4. Results	16
4.1. Experiment 1	16
4.2. Experiment 2	16
5. Conclusions and Future Work	19
5.1. Conclusions	19
5.2. Future Work	19
A. Individual figures of experiment 1	20
B. Individual figures of experiment 2	36

The Auditory Neuroscience Lab at CNS, Boston University



Figure 0.1.: Group picture of the people working in the Auditory Neuroscience Lab at CNS of Boston University. Prof. Barbara G. Shinn-Cunningham can be found in the middle row on the right outermost, myself can be found in the last row on the left outermost.

1. Introduction

Masking in the auditory domain is a well known field in acoustical research. A lot of studies and books about masking exist, especially masking from single sources.

Masking in the 3-dimensional area is a field where in the last few years a lot of research is going on, a lot of them deal with masking virtual auditory space but just a few so far deal with masking and auditory grouping.

If a sound in a complex mixture of different sounds is un audible by the human ear, then the sound is masked by the other components of the complex sound. Generally there are two different types of masking, temporal and spectral masking.

1.1. Spectral (simultaneous) masking

This type of masking describes what happens when two sounds which are located in the same spectral area happens. The louder sound masks the sound, which has a certain level under the level of the louder sound. Sounds in a certain spectral area also mask sounds in neighbouring spectral regions. For a better understanding, please refer to figure 1.1. There you can see masking patterns that are produced with a sine wave with 1kHz center frequency and a certain level. These sine waves suppress not only sounds with the exact same frequency than the masking sound.

1.2. Temporal masking

Temporal masking describes the phenomenon, that maskers also mask sounds, that are some time before and after a masker. Temporal masking is described in figure 1.2. Here you can see the timespan, where sounds from a certain masker are masked before and after that masker.

1.3. Spatial masking

Spatial masking describes the phenomena that sounds are masked from a certain masker, if they are coming from the same direction in space. If they are moved to a certain place that is

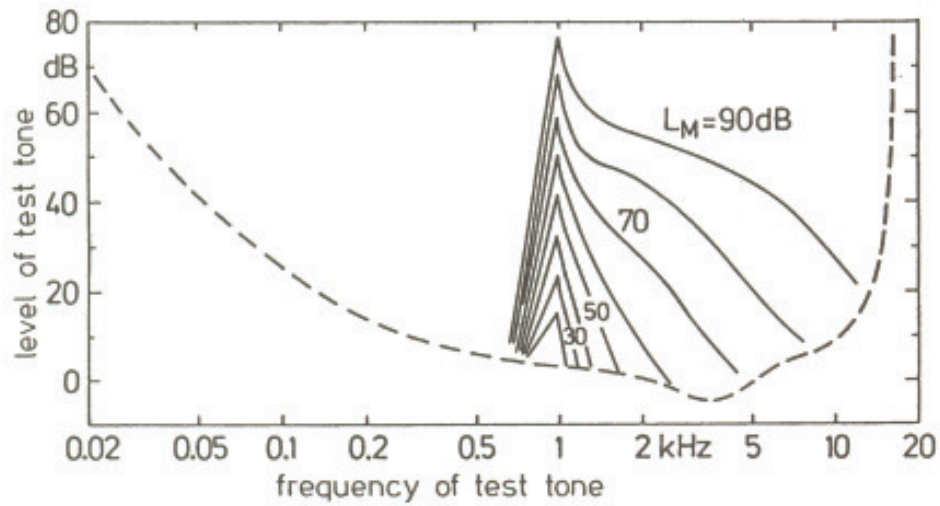


Figure 1.1.: Spectral Masking: Masking patterns produced by a sine wave with 1kHz center frequency at several levels. Picture taken from [2]

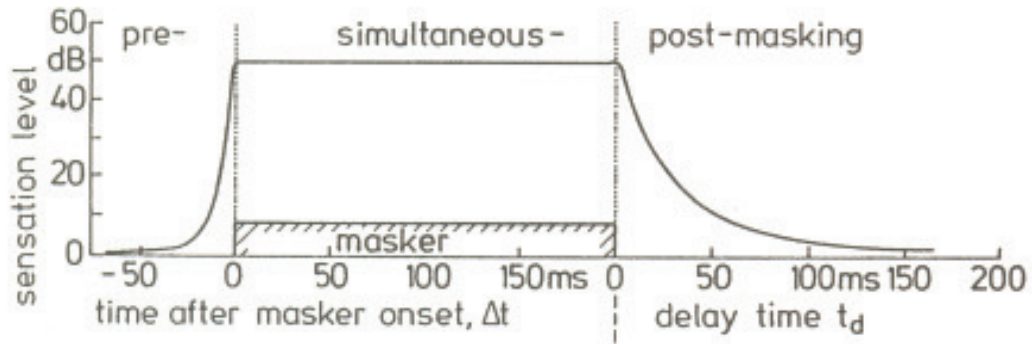


Figure 1.2.: Temporal Masking: Forward, backward and simultaneous masking. Picture taken from [2]

away from the masker, they get unmasked. In this study and in the project we are working on we are looking for the angles, at which spatial unmasking occurs.

2. Methods

In this study we are interested in spatial unmasking. We are presenting sounds from different spatial directions and take a look, if a target sound gets audible again when moving it away from the masker.

To generate a setup to test unmasking in the spatial domain, there are two basic approaches:

1. Sound sources generated with loudspeakers
2. Sound sources generated with headphones

In figure 2.1 the approach with loudspeakers is shown. You can see that in the picture a masker and a maskee (or target) are coming from different directions.

In our study we simply used headphones for placing the sounds in the spatial domain. Both sound sources (masker and target) were placed at the same position in one case at 0° and at about -55° (coming from the right hand side) in the other case. This spatial shifts were generated by applying interaural time differences (ITDs) to the sounds. That means that there is a dime difference between the sound in the right ear and the sound in the left ear. With the help of this time difference humans are able to detect sounds from different locations. In our case 0° and -55° correspond to $0\mu s$ and $-400\mu s$.

2.1. Setup

There were 2 experiments in this study. One where the masker was low frequency and the target was high frequency and one where the masker was high frequency and the target was low frequency. In the experiment where the target was high frequency the masker was a combination tone of 250Hz and 750Hz sinusoids and the target was a combination tone of 3750Hz and 4259Hz sinusoids, both combination tones comodulated with a 250Hz modulation rate SAM. The tones were applied with 10ms raised cosine ramps at the onset and the offset. In the second condition the masker was an amplitude modulated combination tone of 3750Hz and 4250Hz sinusoids and the target a amplitude modulated combination tone of 250Hz and 750Hz sinusoids like described earlier but the other way round.

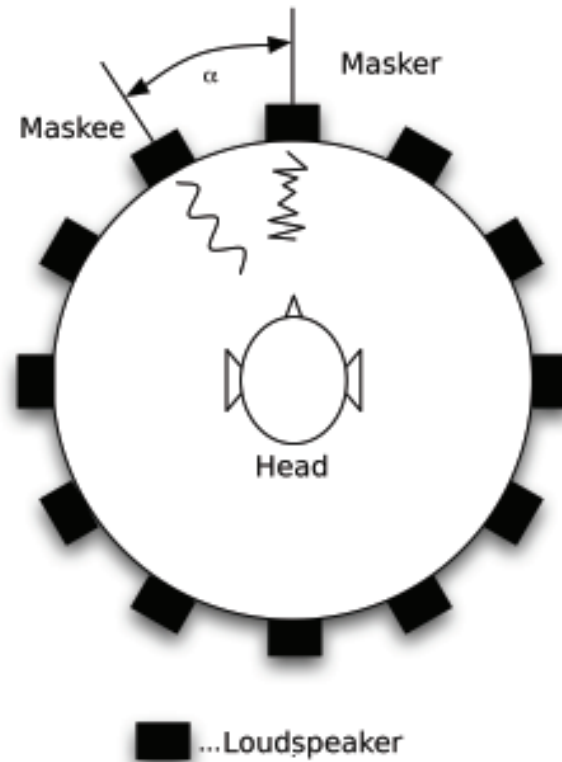


Figure 2.1.: Test setup for presenting a masker and a maskee (or target) from different spatial locations. In this picture one can see a setup with discrete loudspeakers, this can also be simulated using a virtual auditory space.

For experiment 1 a bandlimited pink noise signal with a low cutoff frequency of 100Hz and a high cutoff frequency of 1000Hz was generated. The energy in the 1/3rd octave bandwidth centered at 500Hz was adjusted to be the same than the energy of the low-frequency combination tone. For experiment 2 a bandlimited pink noise signal between 2000Hz and 6000Hz was generated, again the energy in the third octave bandwidth centered at the center frequency of the combination tone (in this case 4000Hz) was adjusted to be the same than the energy of the amplitude modulated high frequency combination tone.

In each of the two experiments there were basically three conditions:

1. Single: Masker and Target presented once
2. Streamed: The Masker presented 4 times preceding Masker and Target together
3. Control: A bandlimited noise signal preceding Masker and Target together

Masker and Target were always presented in time slots. The length of a time slot was 250ms and the length of the pause between the time slots was 50ms. In the single condition Target and Masker were played at the same time lasting 250ms. In the streamed condition four 250ms

timeslots of the Masker were presented preceding the Target and the Masker presented at the same time slot and in the control condition a continuous bandlimited noise was played instead of the masker and after a 50ms pause Masker and Target were played within 1 time slot. For a better explanation of the test setup of experiment 1 and experiment 2 see chapter 3.1 and chapter 3.2 respectively.

Each of these conditions always were played with 2 different ITDs, namely $0\mu\text{s}$ and $-400\mu\text{s}$.

All sounds were generated offline with MATLAB 2009a. 12 different representations of each of the two bandlimited pink noise signals were generated. At playback time one of the 12 representations was picked randomly to provide a better random noise distribution. As playback device a Tucker-Davies Technology hardware with a sampling frequency of 25kHz was used for D/A conversion. 0dB in the results map to 80dB(A) measured with a Bruel Kjaer calibrator and frequency band analyzer.

All experiments were presented in a sound treated booth with closed Sennheiser HD 280 Pro Headphones.

2.2. Participants

12 subject participated in the experiment, whereas 6 subjects (3 female, 3 male; ages between 18 and 46 years) only listened to the high frequency target and low frequency masker and 6 subjects (3 female, 3 male; ages between 18 and 33 years) only listened to the low frequency target and high frequency masker. An informed consent was given by all participants. The subjects had to do an initial hearing screening and they had to have a good hearing with a hearing loss not greater than 20dB.

2.3. Procedure

Participants were seated in a sound-treated booth in front of a PC monitor and a PC mouse. The controlling of the experiment by the participants was done using a Graphical User Interface in MATLAB. At first they had to input their name into the specified text field in the graphical user interface, see figure 2.2.

At the beginning of each trial, the participants could hear the target separated as often as they wanted to, see figure 2.3.

The thresholds were determined by an adaptive 2 Alternative Forced Choice task (see figure 2.4). The 2-AFC task was combined with a 2up, 1down staircase method. The algorithm always decreased the masker level in a way, that in the first 2 returns the step size was 5 db, for the next 4 returns it was 3 db and in the end (for the last 6 returns) it was decreased to 1 dB.



Figure 2.2.: Beginning of the listening test. The participants had to input their name.



Figure 2.3.: The Target was presented to the participants. They could listen to it as often as they wanted to.

Hereby only the last 6 data points were taken to calculate the threshold. Furthermore only thresholds where the variance of the mean of the last 6 data points was less than 8dB were taken into account for the results. Please have a look at the appendix to see the individual

tracks of all the subjects that participated in the study.

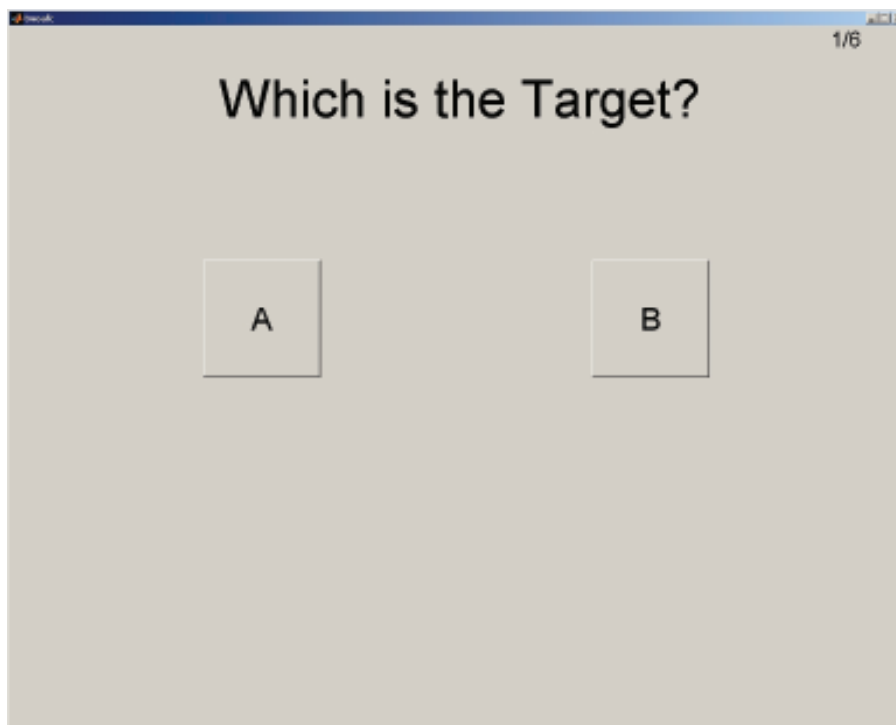


Figure 2.4.: GUI of the 2-alternative forced choice task presented in the listening test.

The subjects got a feedback, whether they were right (see figure 2.5) or wrong (see figure 2.6).

The different trials were presented randomly as well as the different conditions (single collocated/separated, streamed collocated/separated, noise collocated/separated) were presented randomly. Each participant had to do each trial 4 times, to get a more stable threshold. The stimulus with the target was randomly presented on position A or B.

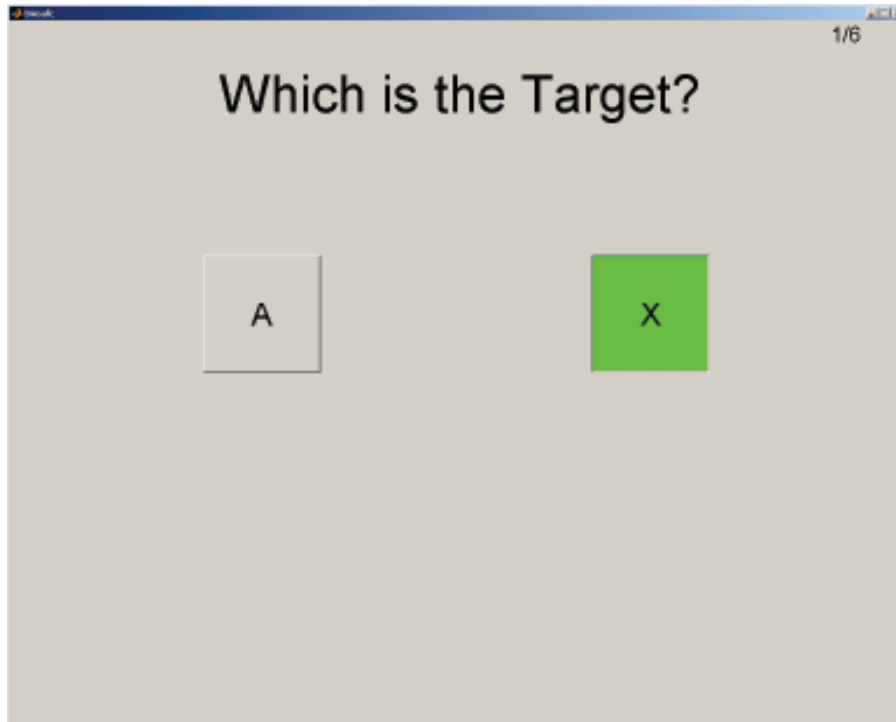


Figure 2.5.: If the participant hit the target, the button flashed green.

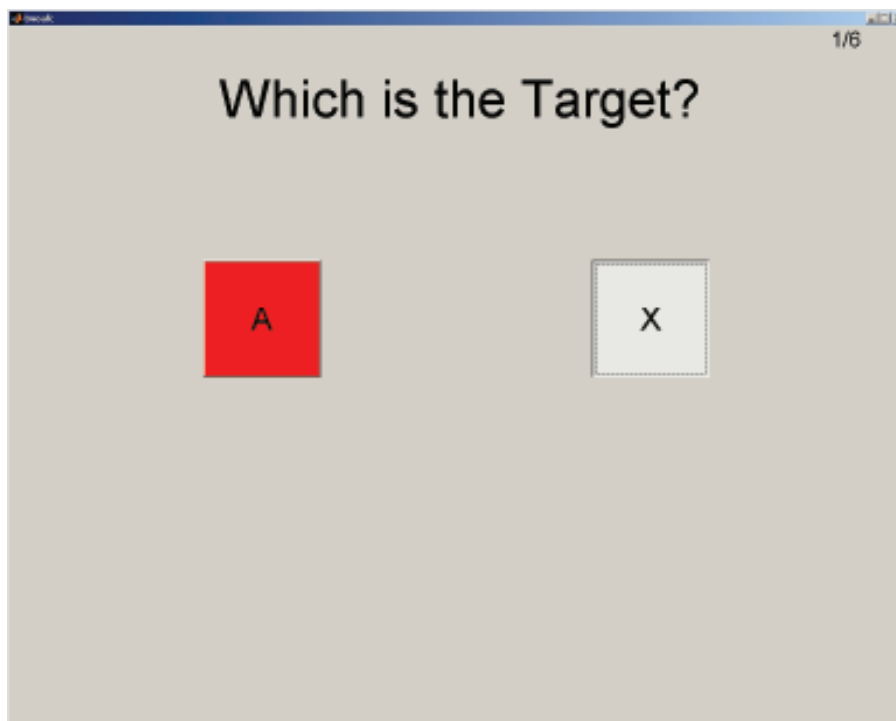


Figure 2.6.: If the participant missed the target, the unpressed button flashed red.

3. Experiments

3.1. Experiment 1

As adumbrated in section 2.1 participant were presented three different conditions, each of them with 2 different ITDs. The participants had to do a 2-alternative forced choice task where they were asked to identify the sound example, which contains the Target. In figure 3.1 one can see the three different conditions. Each of the blocks represents a sound.

The participants had to do 4 repetitions of each trial run, whereas each trial run took them about 4 minutes. Participants had to do 6 different trial runs at once (Single collocated/separated, Streamed collocated/separated and Control collocated/separated), that means that one run was about 24 minutes long. So the whole test procedure was divided into 2 sessions of about 48 minutes, which had to be on different days. One session again was split into two parts with a short break inbetween.

3.2. Experiment 2

In experiment 2 again 3 different conditions were presented to the participants, which varied in their ITD (collocated $0\mu s$, separated $-400\mu s$). They had to do a 2-alternative forced choice task where they had to find the sound example, where the Target was involved in. Please refer to figure 3.2 for a detailed sketch of the three conditions.

The participants had to do 4 repetitions of each trial run, whereas each trial run took them about 5 minutes. Participants had to do 6 different trial runs at once (Single collocated/separated, Streamed collocated/separated and Control collocated/separated), that means that one run was about 30 minutes long. So the whole test procedure was divided into 2 sessions of about 60 minutes. One session again was split into two parts with a short break inbetween, like in experiment 1.

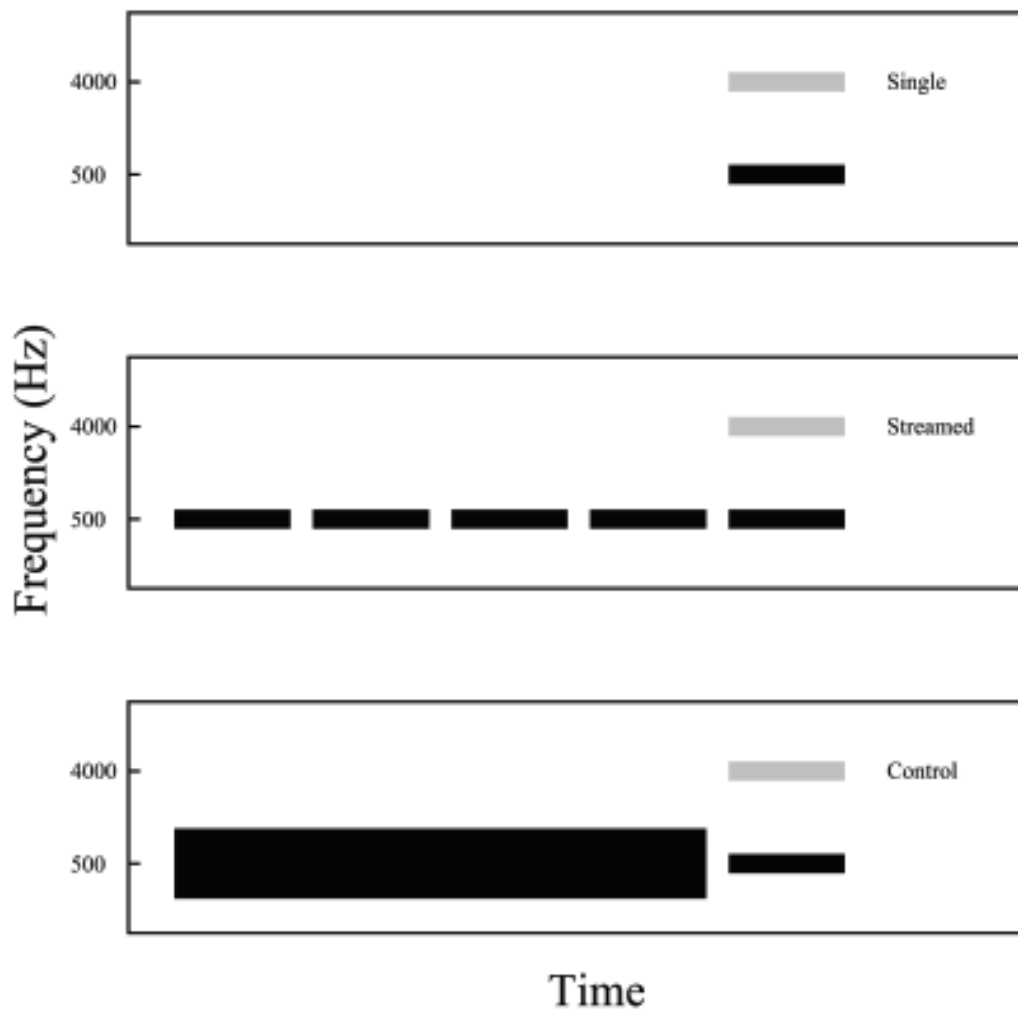


Figure 3.1.: The three conditions for experiment 1. Participants were presented these 3 different types of stimuli with 2 different ITDs.

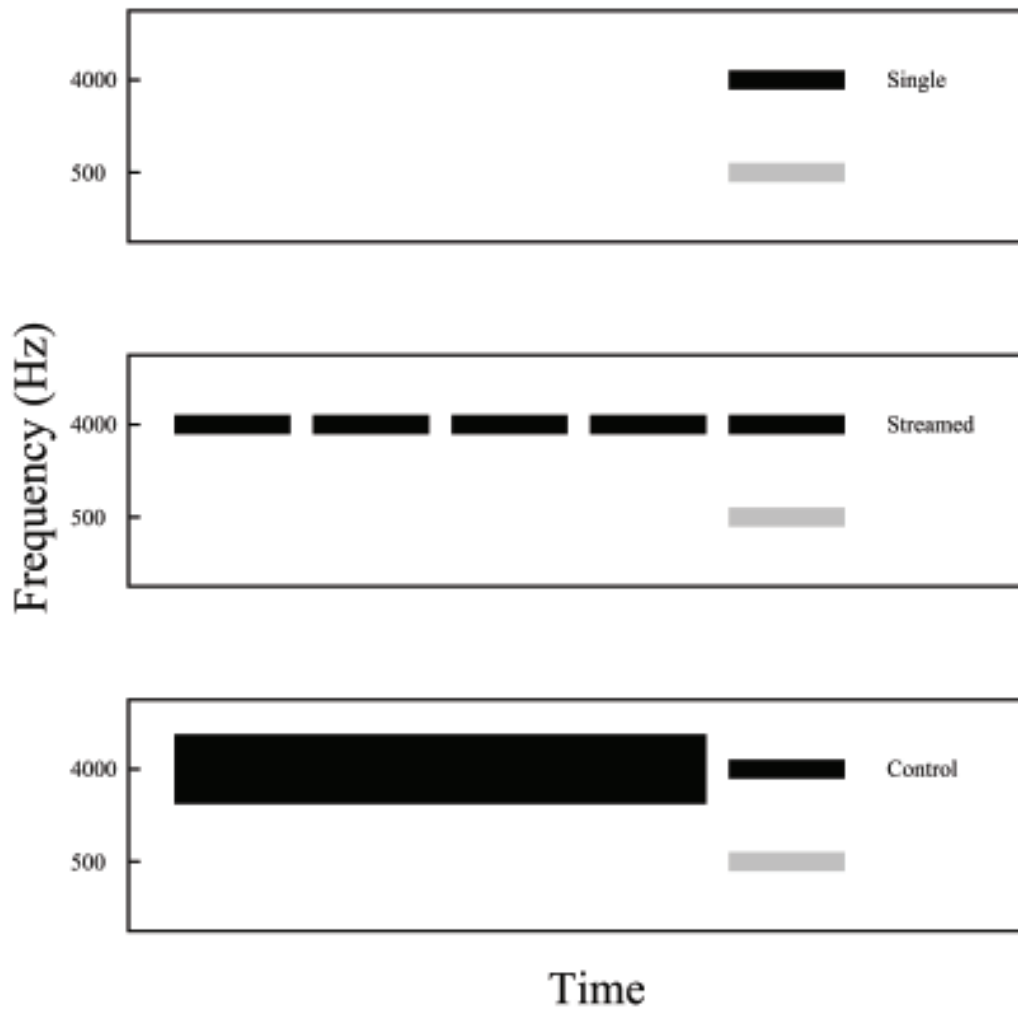


Figure 3.2.: The three conditions for experiment 2. Participants were presented these 3 different types of stimuli with 2 different ITDs.

4. Results

4.1. Experiment 1

The results of experiment 1 are sketched in figure 4.1.

What we can see here is the threshold, where the participants could just hear the target in presence of the masker. 0dB means, the target has the same level as the masker.

We can observe, that for the condition 'Streamed' the masked threshold is lower than for the 'Single' and 'Control' condition. This is, because the four representations of the masker, which precede the masker and the target, are summarized to one auditory stream in the human auditory unit. Therefore the human ear is more sensitive to new sounds which follow this auditory stream¹.

We can also observe that there is spatial unmasking in every case. Obviously the data here is not statistically significant, but the given results show a promising tendency. To be sure that the data is relevant and to get significant results hopefully some more participants have to be tested.

The result figures of every single participant are shown in appendix A.

4.2. Experiment 2

The results of experiment 2 are presented in figure 4.2. Generally speaking participants took a bit more time for completing experiment 2. This is, because the masked threshold was much lower than in experiment 3. Also the masked thresholds of all conditions and ITD-variations are very close together. This might be, because no masking takes place in experiment 2. This can also be seen when looking at the masking patterns in [2]. That means that no unmasking effects or streaming effects can be observed here, the experiment just showed that there is no masking with an amplitude modulated combination tone of 3750Hz and 4250Hz sinusoids as masker and an amplitude modulated combination tone of 250Hz and 750Hz sinusoids as target.

The individual result figures of the participants can be found in appendix B.

¹This auditory stream which is summarized to a unit is called an auditory object.

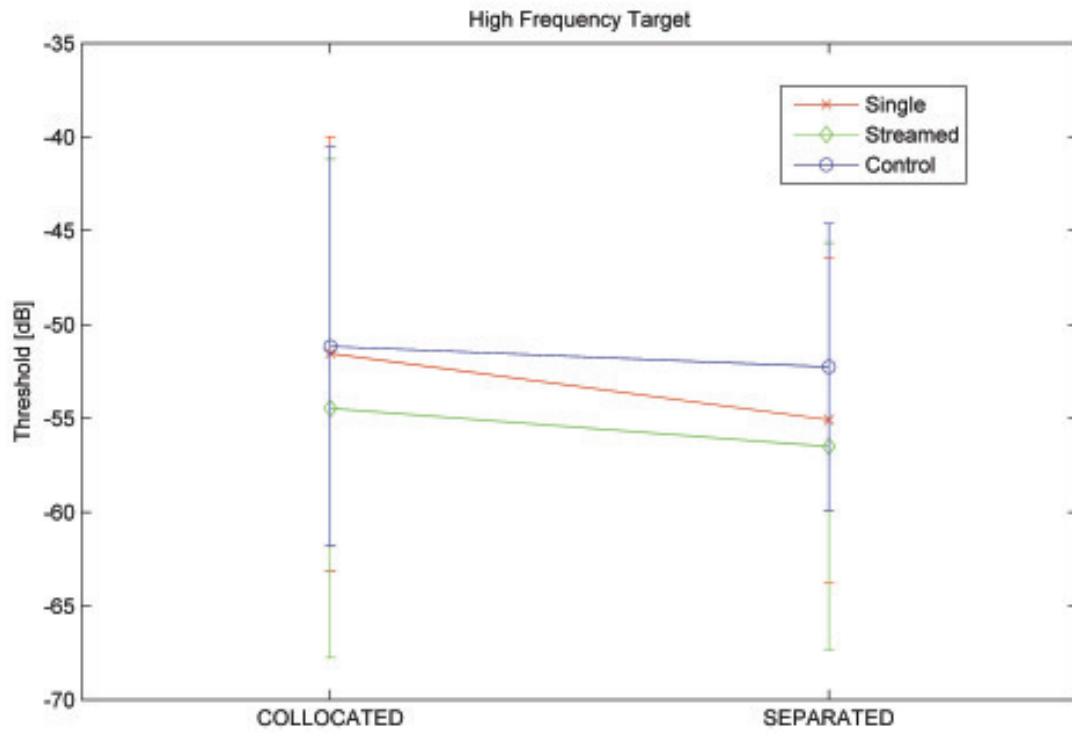


Figure 4.1.: Results of experiment 1. On the left hand side there are the results for the masker and the target presented with the same ITD (collocated) and on the right hand side there are the results where the target has an ITD of $-400\mu s$. The red line with the crosses represents the Single condition, the green line with the diamonds represents the Streamed condition and the blue line with the circles represents the Control condition.

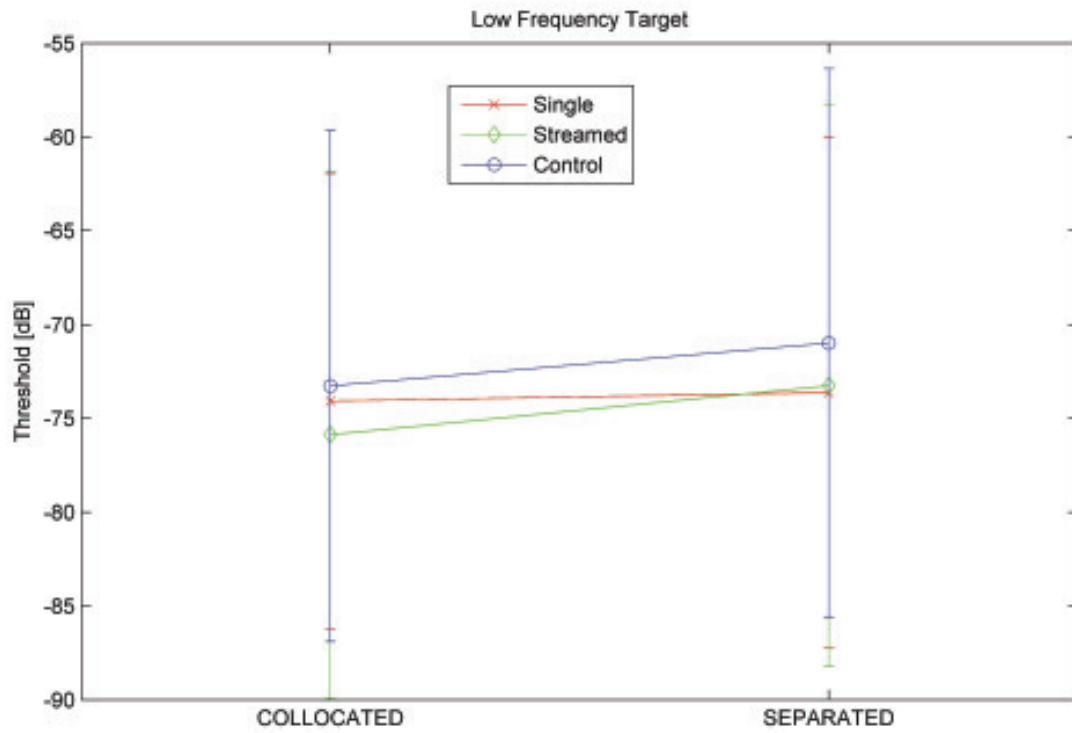


Figure 4.2.: Results of experiment 2. On the left hand side there are the results for the masker and the target presented with the same ITD (collocated) and on the right hand side there are the results where the target has an ITD of $-400\mu\text{s}$. The red line with the crosses represents the Single condition, the green line with the diamonds represents the Streamed condition and the blue line with the circles represents the Control condition.

5. Conclusions and Future Work

5.1. Conclusions

We have introduced a study about spatial masking and streaming with two different masker and target setups. The interaural time differences of the target had two different values, $0\mu s$ in one case and $-400\mu s$ (coming from the right) in the other case. All experiments in the study were performed using headphones.

We could see, that in the condition where the masker was at a higher frequency than the target, there was no masking at all, participants could hear the target until the level of the target went below their hearing threshold.

In the case where the frequency of the target was higher than the frequency of the masker, we could see that masking occurred and so did spatial release of masking. Furthermore we could show that the masking threshold in the 'Streamed' condition is lower than in the other conditions.

Unfortunately the data has a big variance and is not really statistically significant, therefore some additional subjects have to be tested.

5.2. Future Work

This experiment can be expanded in many directions. First other stimuli can be used, for example a modulated noise or broadband white noise bursts. Secondly more ITDs can be added to the signals to test the whole horizontal plane masking and streaming effects. Then interaural level differences (ILD) can also be taken into account.

Right now 8 additional subjects are tested at Boston University to gain more data. The test setup is identical with the setup in Experiment 1, except 1 additional parameter is asked, namely the absolute hearing threshold of the actual target. This is to check if the masked threshold lies above the hearing threshold.

A. Individual figures of experiment 1

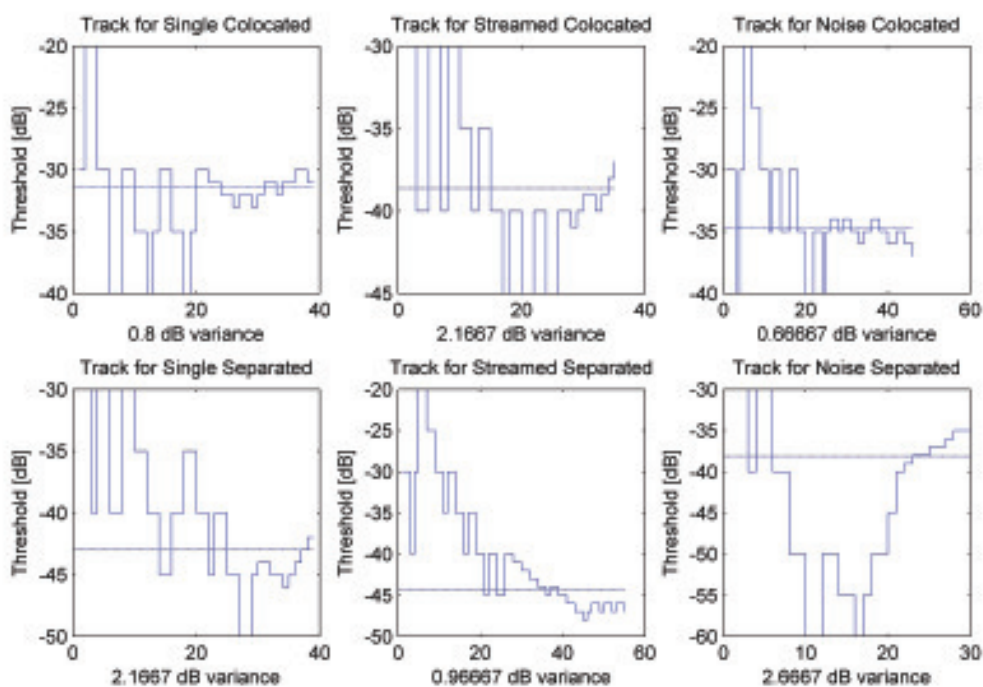


Figure A.1.: Tracks of the first session of subject 1. The dash-dotted line represents the actual masked threshold. The solid line represents the actual threshold of the target which is presented to the participant. Note that only the 6 last reversals are taken into the calculation of the masking threshold.

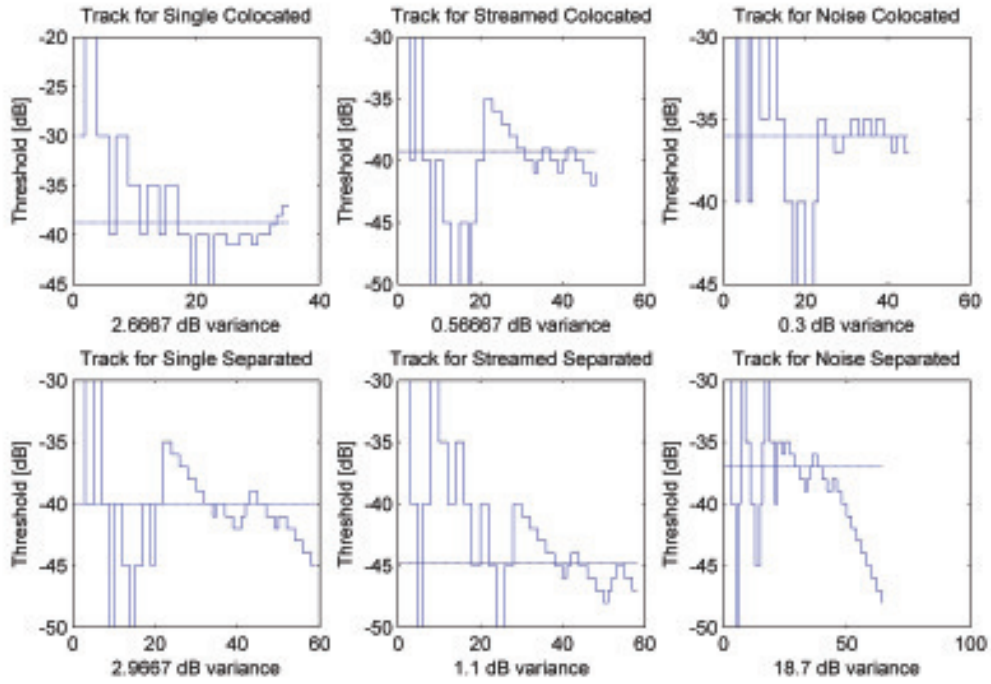


Figure A.2.: Tracks of the second session of subject 1. For details please refer to figure A.1.

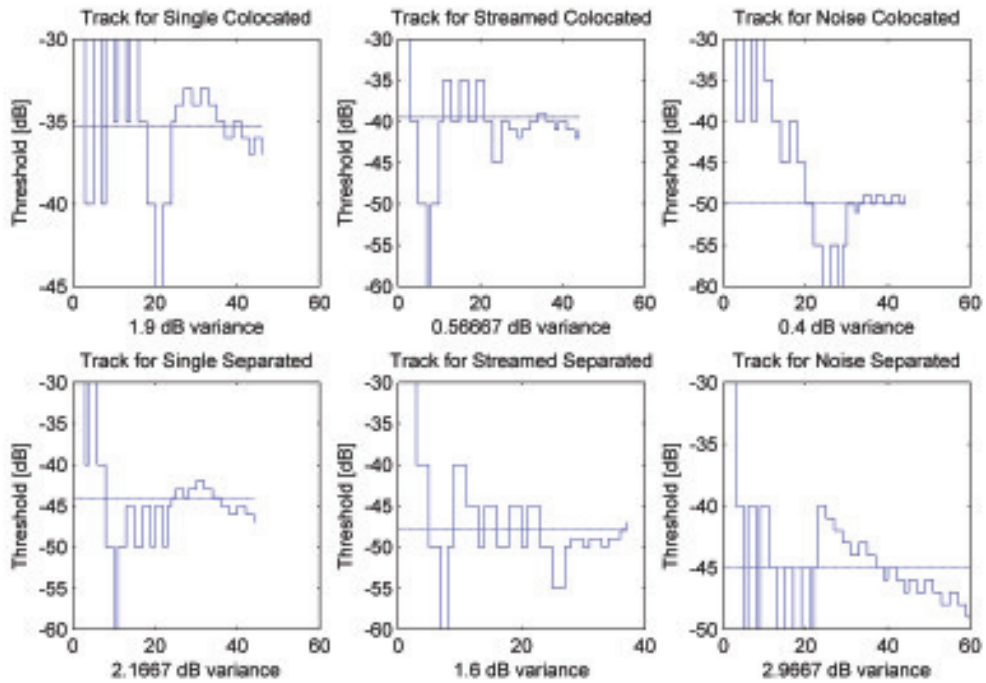


Figure A.3.: Tracks of the third session of subject 1. For details please refer to figure A.1.

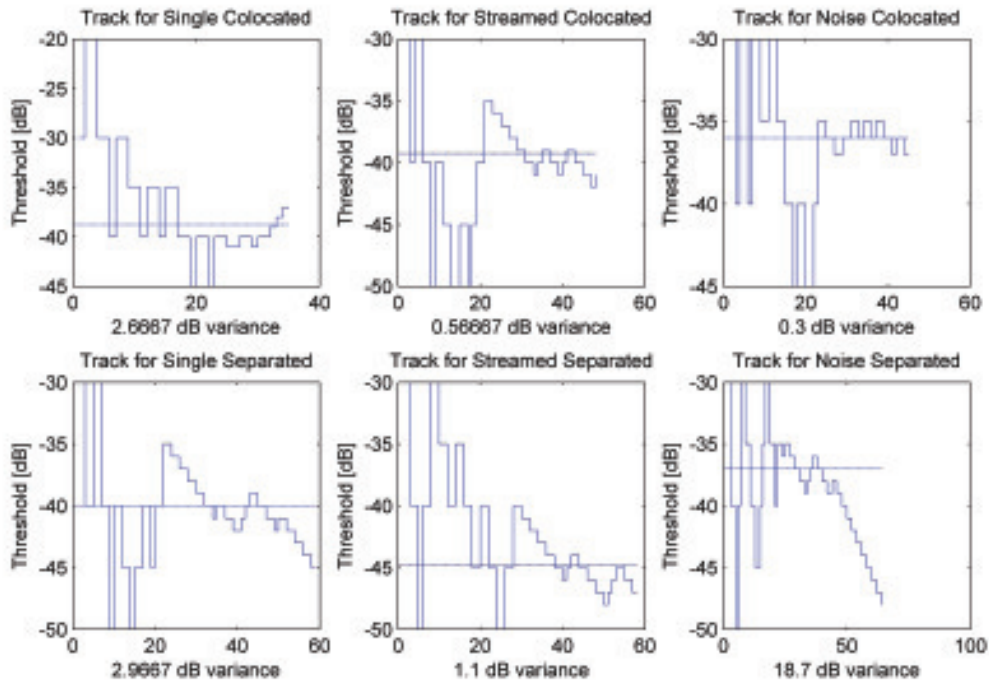


Figure A.4.: Tracks of the fourth session of subject 1. For details please refer to figure A.1.

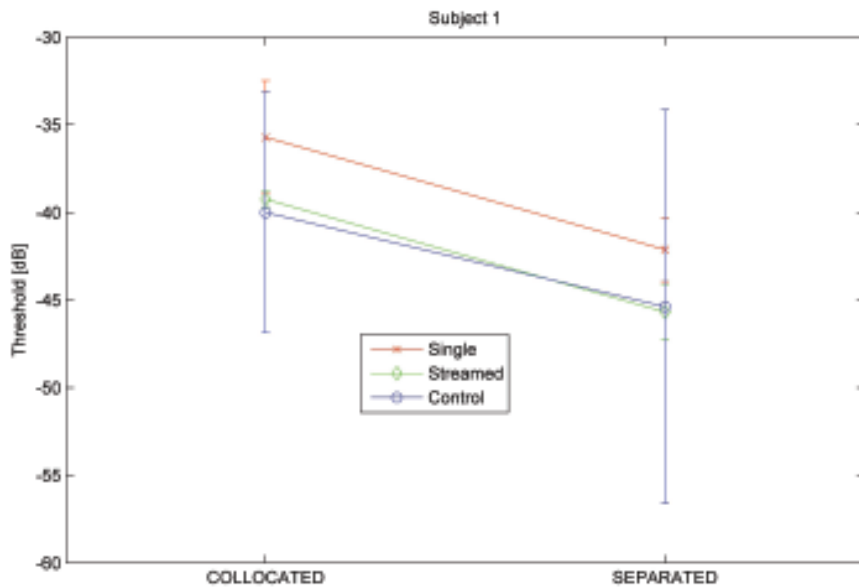


Figure A.5.: Results of subject 1. On the left hand side there are the results for the masker and the target presented with the same ITD (collocated) and on the right hand side there are the results where the target has an ITD of $-400\mu s$. The red line with the crosses represents the Single condition, the green line with the diamonds represents the Streamed condition and the blue line with the circles represents the Control condition.

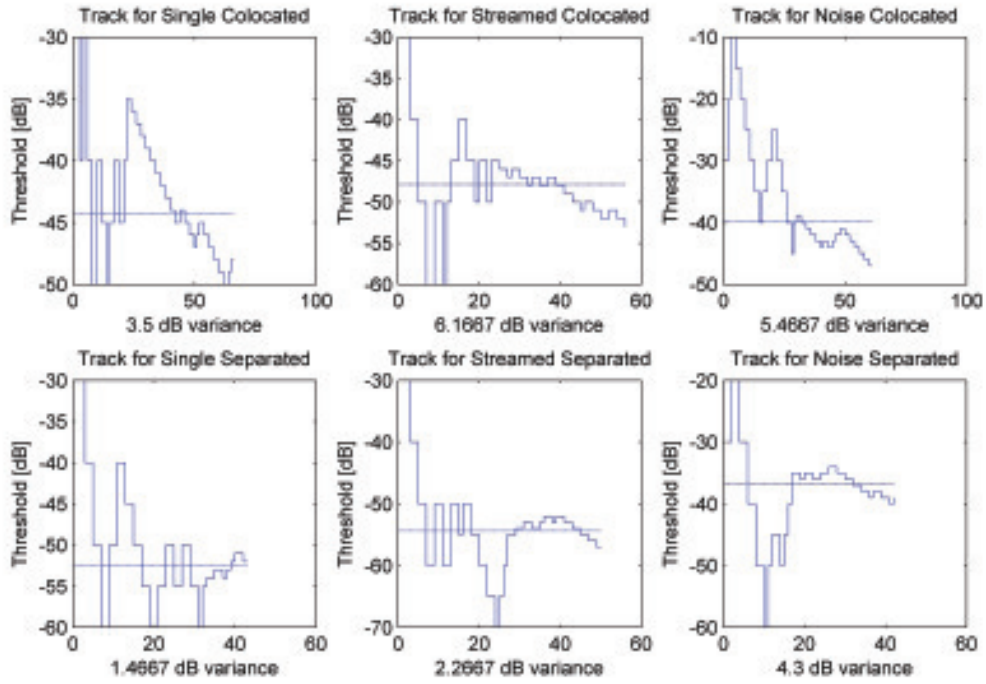


Figure A.6.: Tracks of the first session of subject 2. For details please refer to figure A.1.

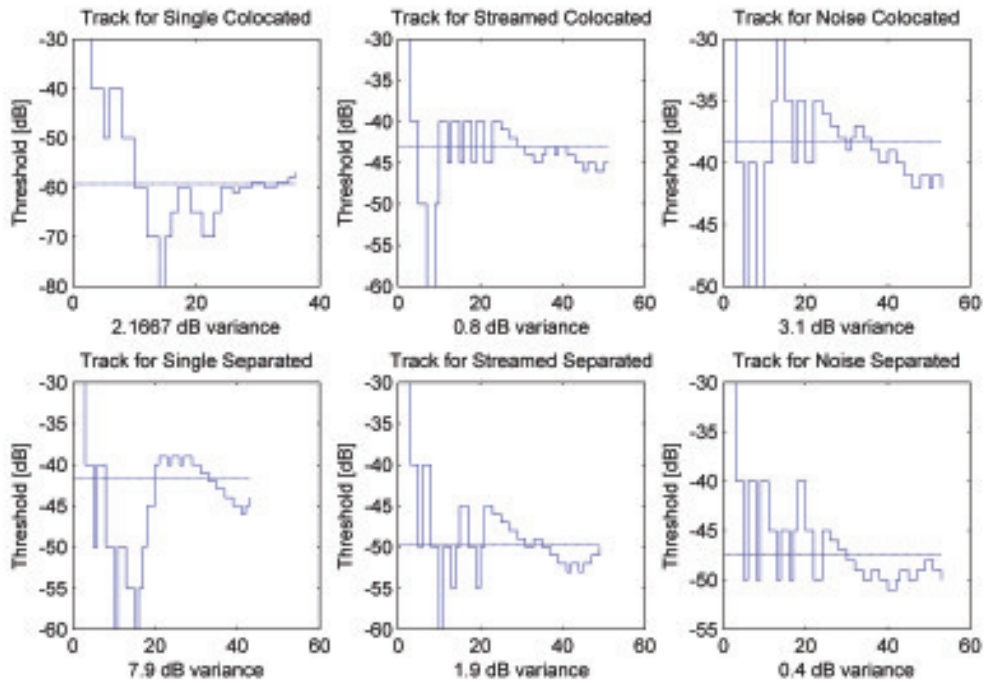


Figure A.7.: Tracks of the second session of subject 2. For details please refer to figure A.1

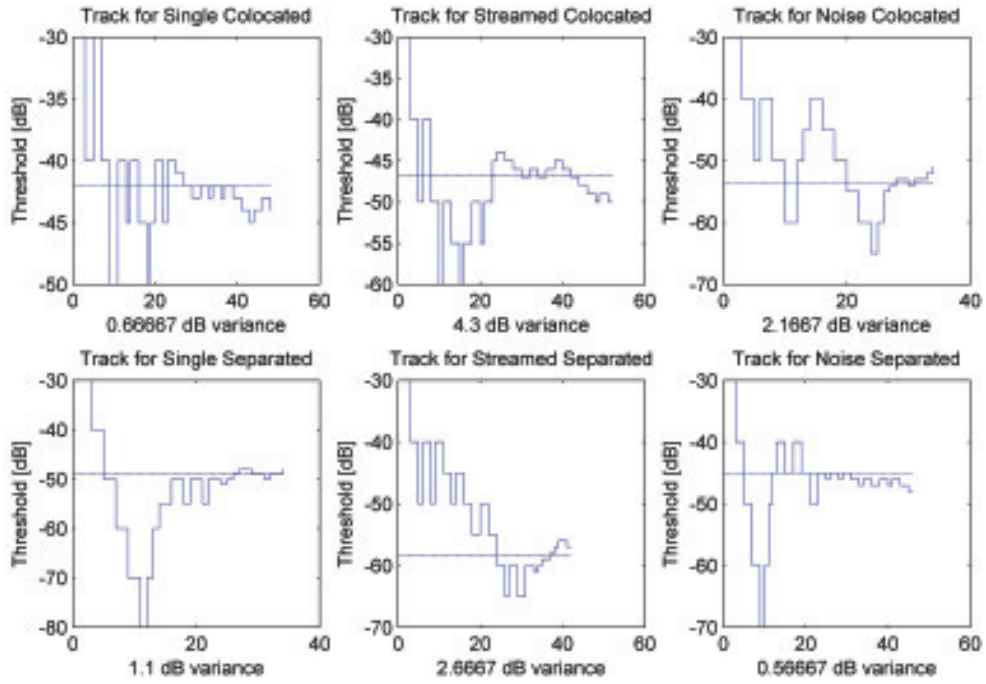


Figure A.8.: Tracks of the third session of subject 2. For details please refer to figure A.1

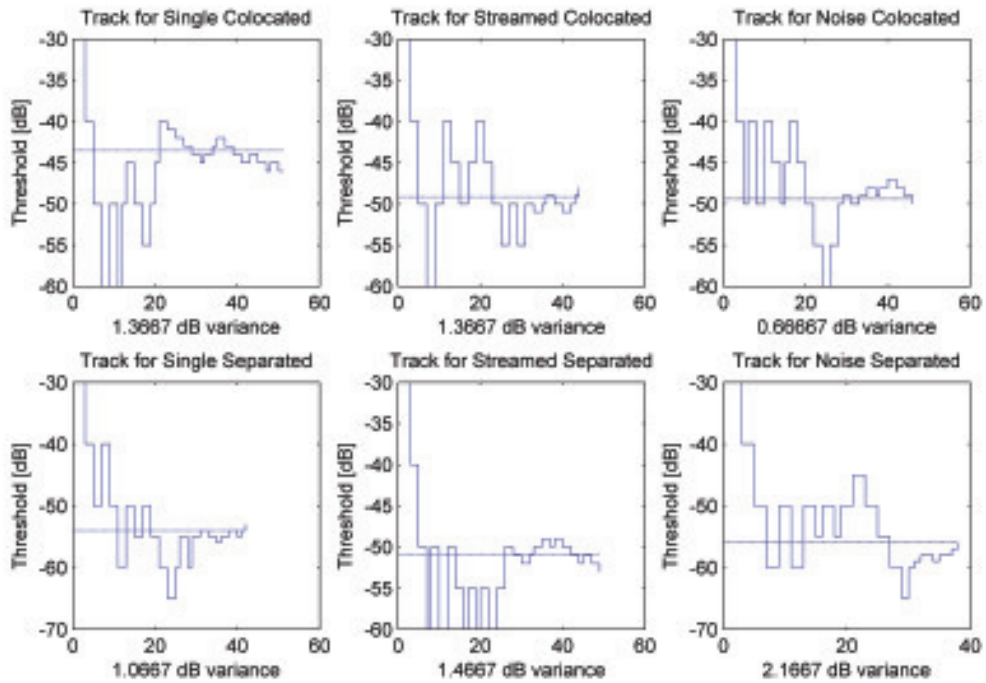


Figure A.9.: Tracks of the fourth session of subject 2. For details please refer to figure A.1

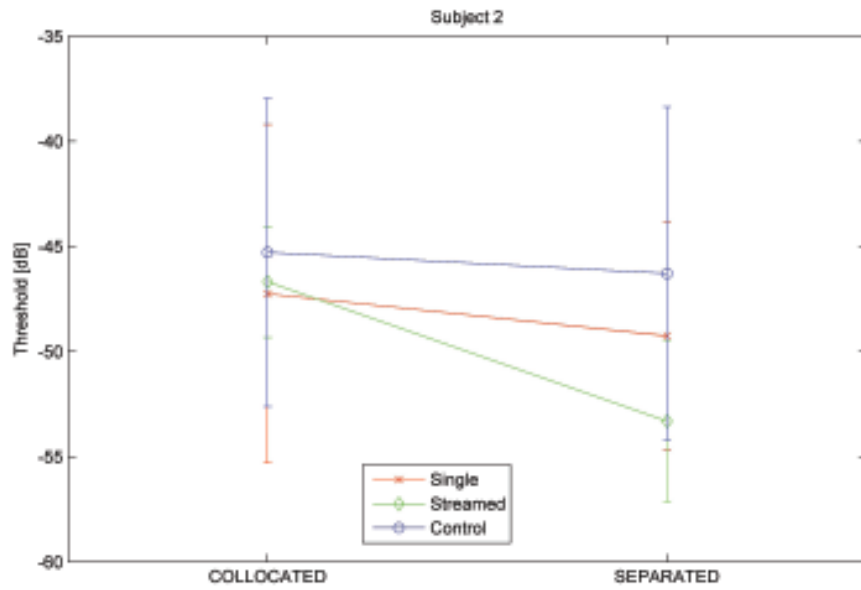


Figure A.10.: Results of subject 2. Please refer to figure A.6 for a description.

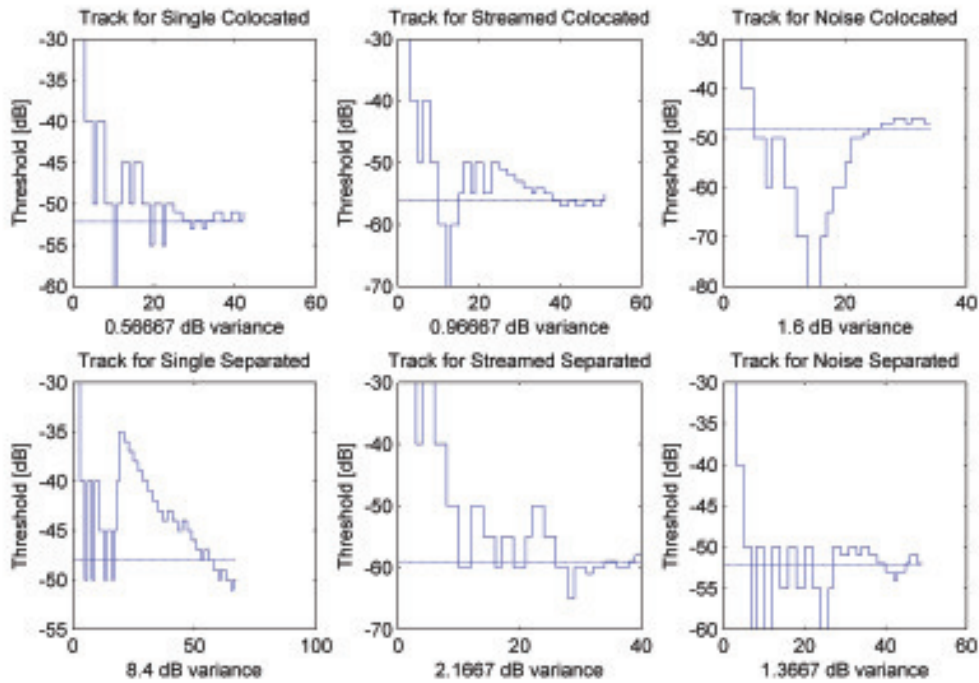


Figure A.11.: Tracks of the first session of subject 3. For details please refer to figure A.1.

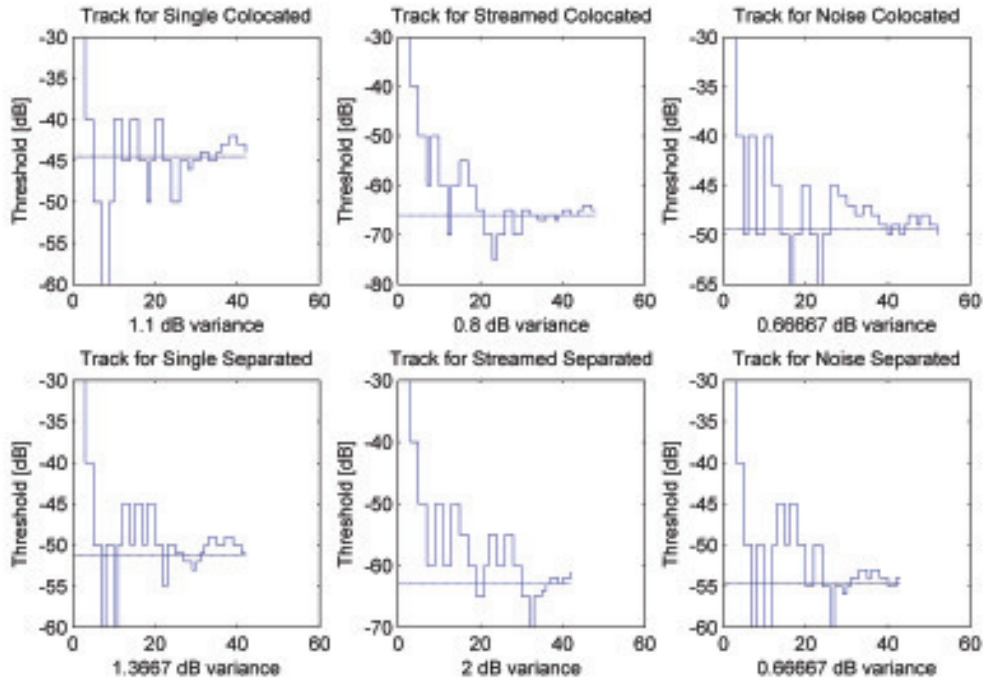


Figure A.12.: Tracks of the second session of subject 3. For details please refer to figure A.1.

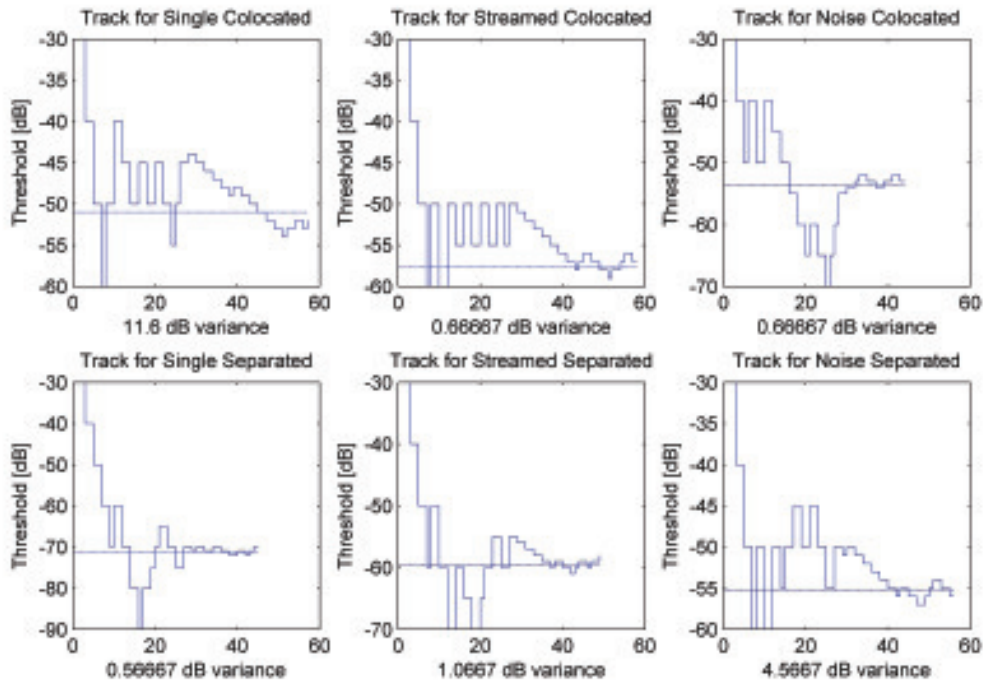


Figure A.13.: Tracks of the third session of subject 3. For details please refer to figure A.1.

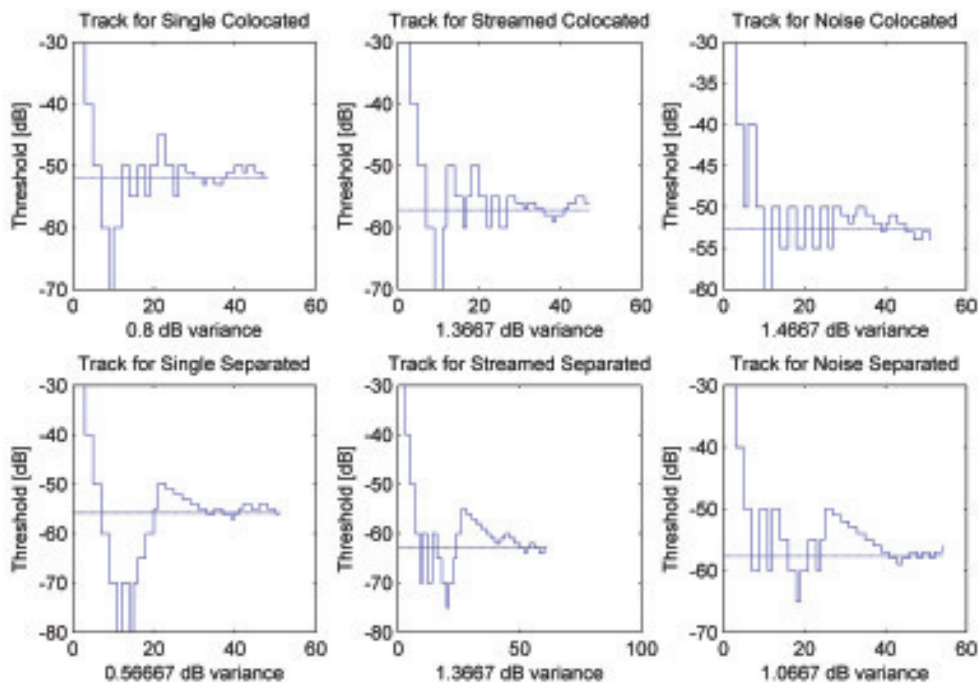


Figure A.14.: Tracks of the fourth session of subject 3. For details please refer to figure A.1.

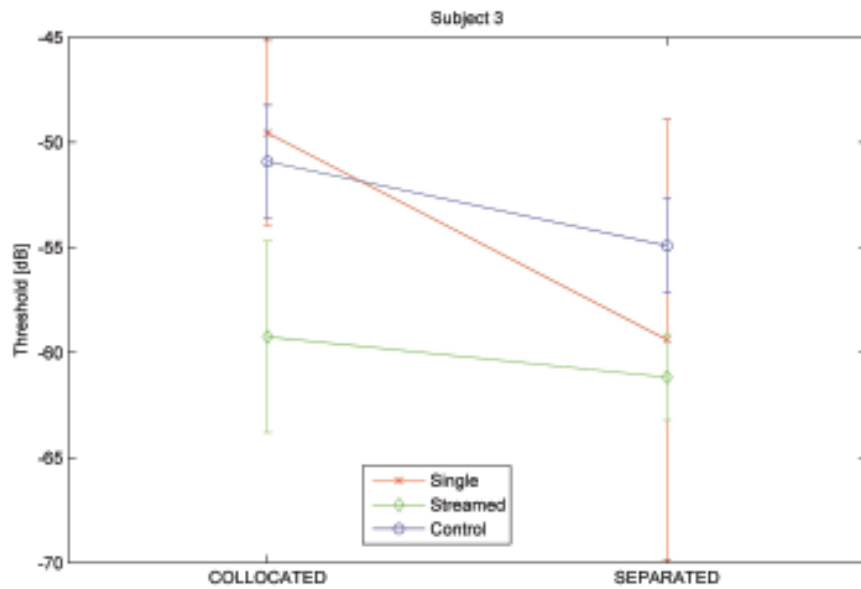


Figure A.15.: Results of subject 3. Please refer to figure A.6 for a description.

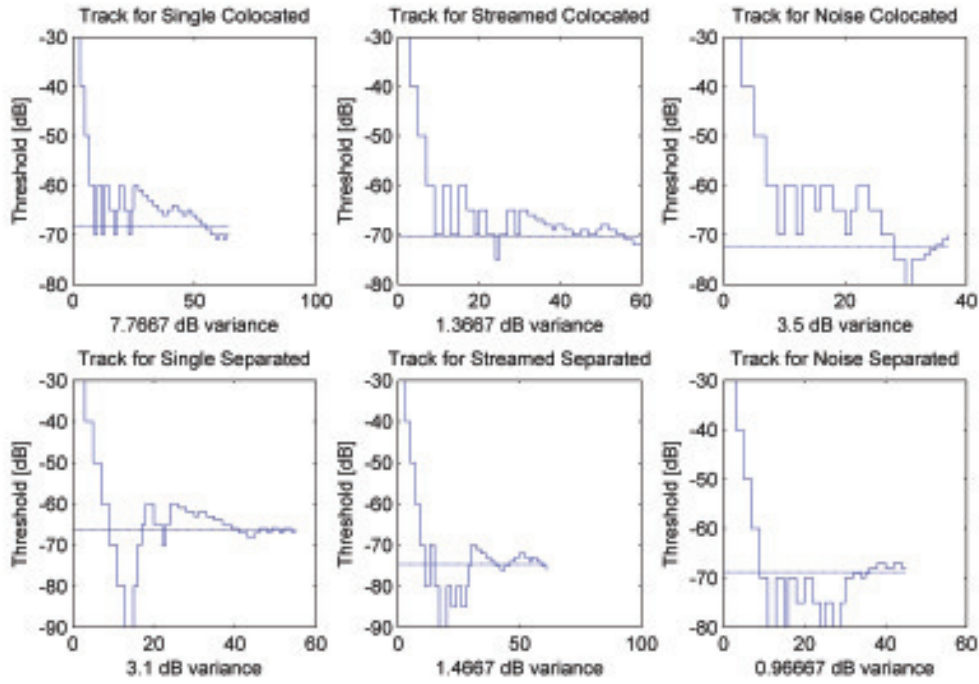


Figure A.16.: Tracks of the first session of subject 4. For details please refer to figure A.1.

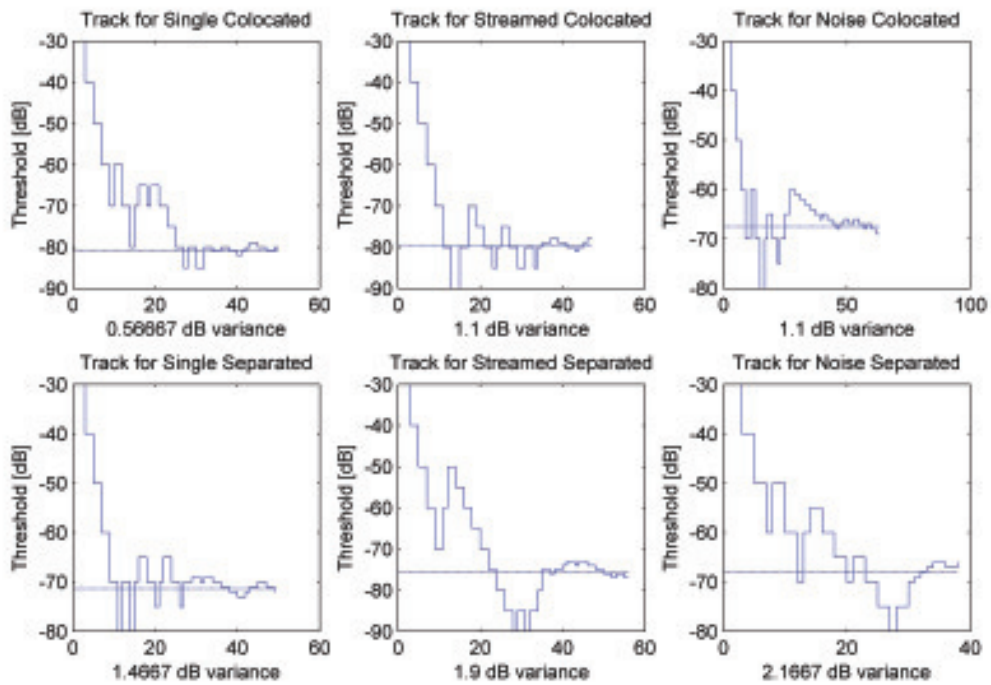


Figure A.17.: Tracks of the second session of subject 4. For details please refer to figure A.1.

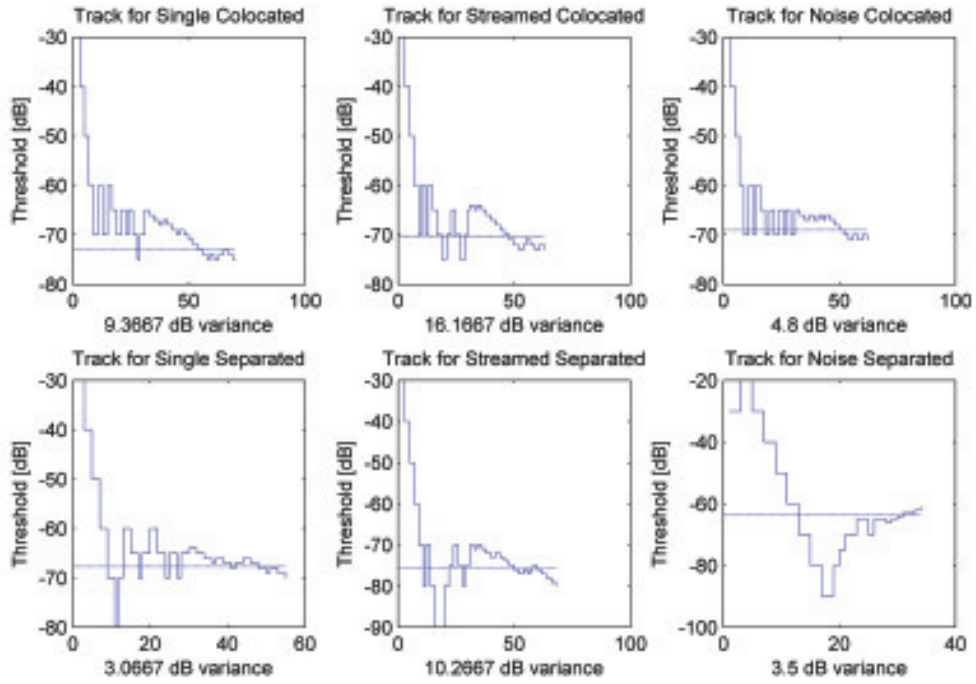


Figure A.18.: Tracks of the third session of subject 4. For details please refer to figure A.1.

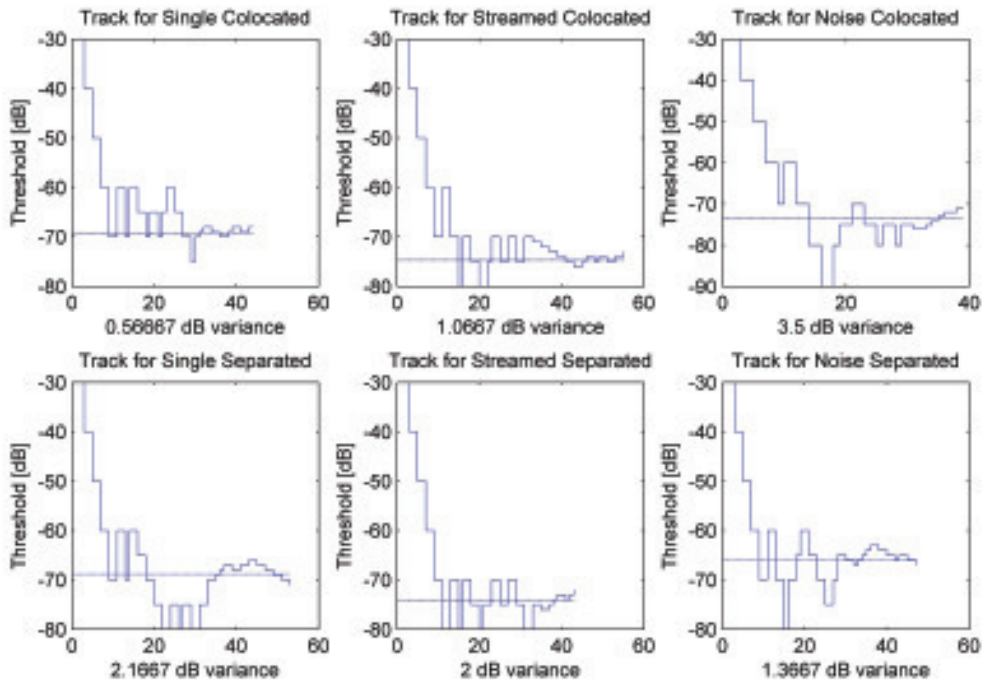


Figure A.19.: Tracks of the fourth session of subject 4. For details please refer to figure A.1.

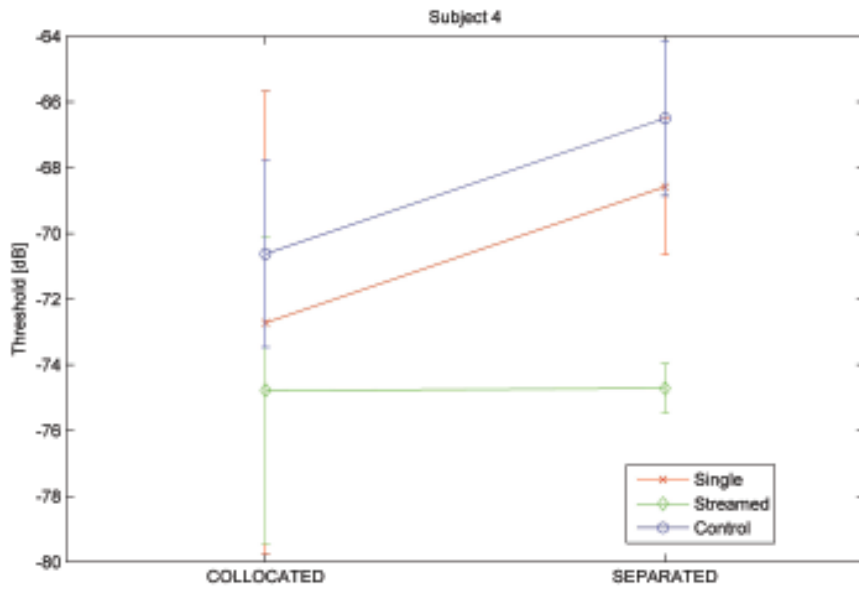


Figure A.20.: Results of subject 4. Please refer to figure A.6 for a description.

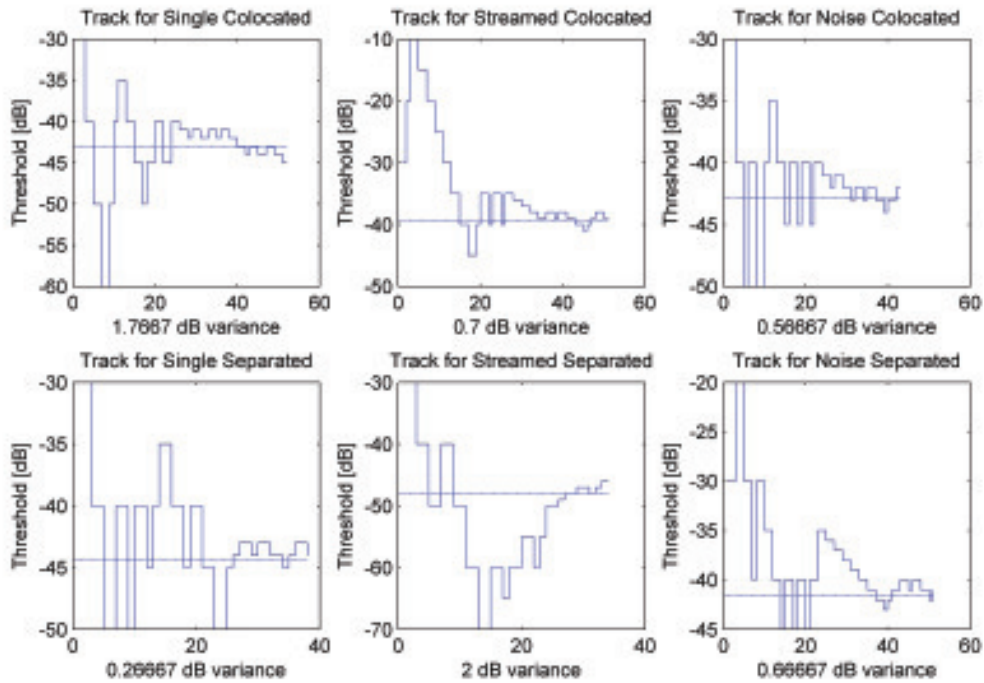


Figure A.21.: Tracks of the first session of subject 5. For details please refer to figure A.1.

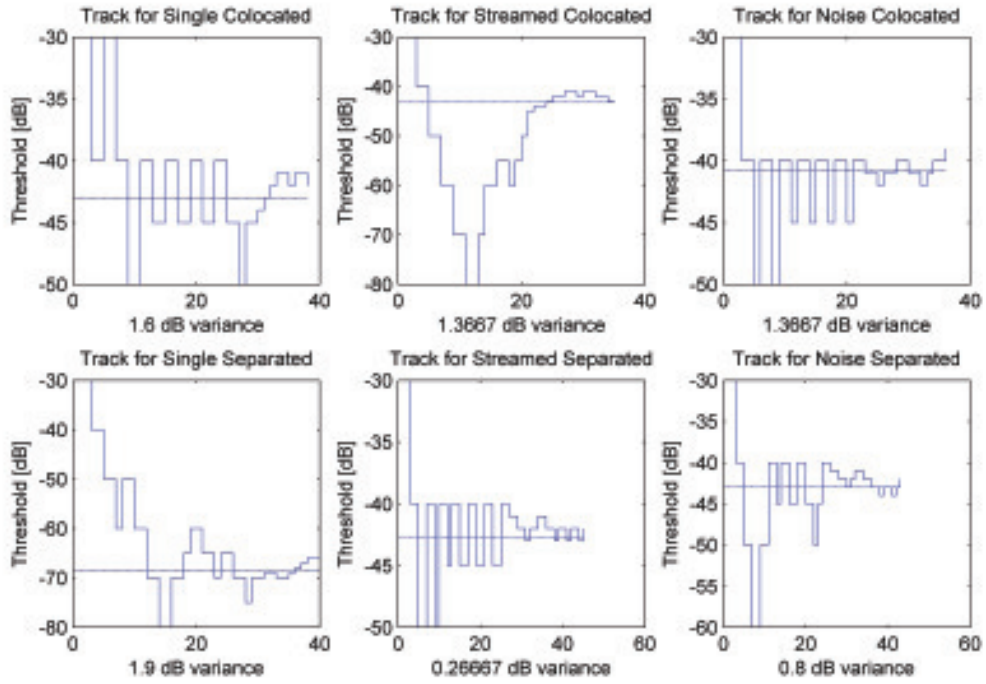


Figure A.22.: Tracks of the second session of subject 5. For details please refer to figure A.1.

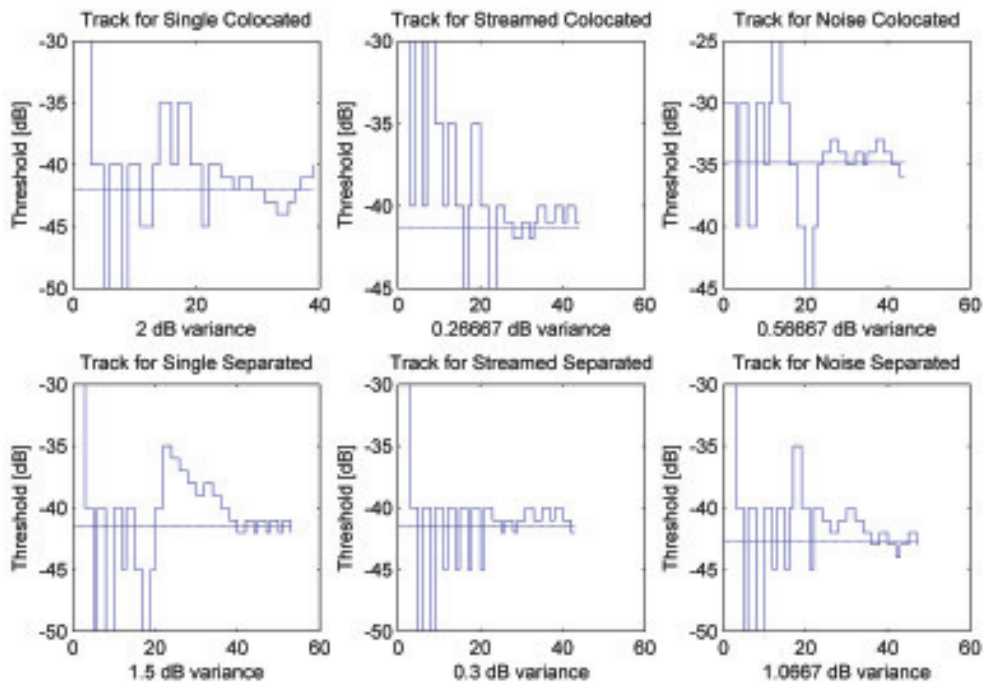


Figure A.23.: Tracks of the third session of subject 5. For details please refer to figure A.1.

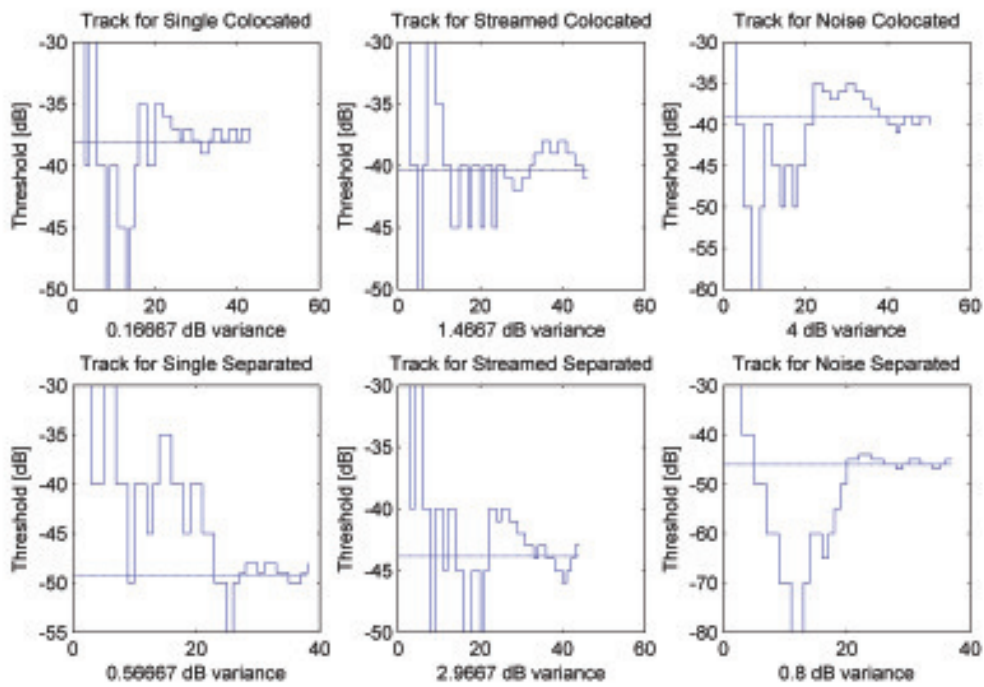


Figure A.24.: Tracks of the fourth session of subject 5. For details please refer to figure A.1.

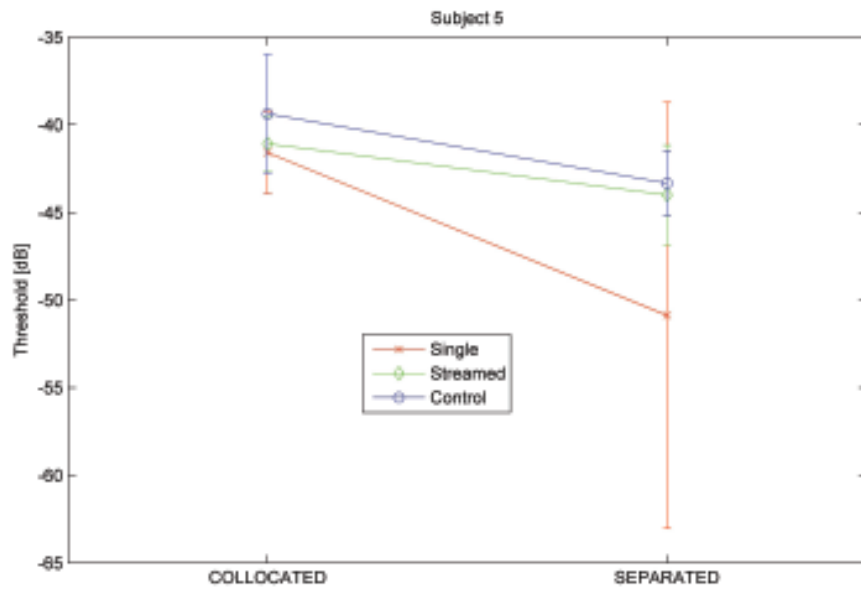


Figure A.25.: Results of subject 5. Please refer to figure A.6 for a description.

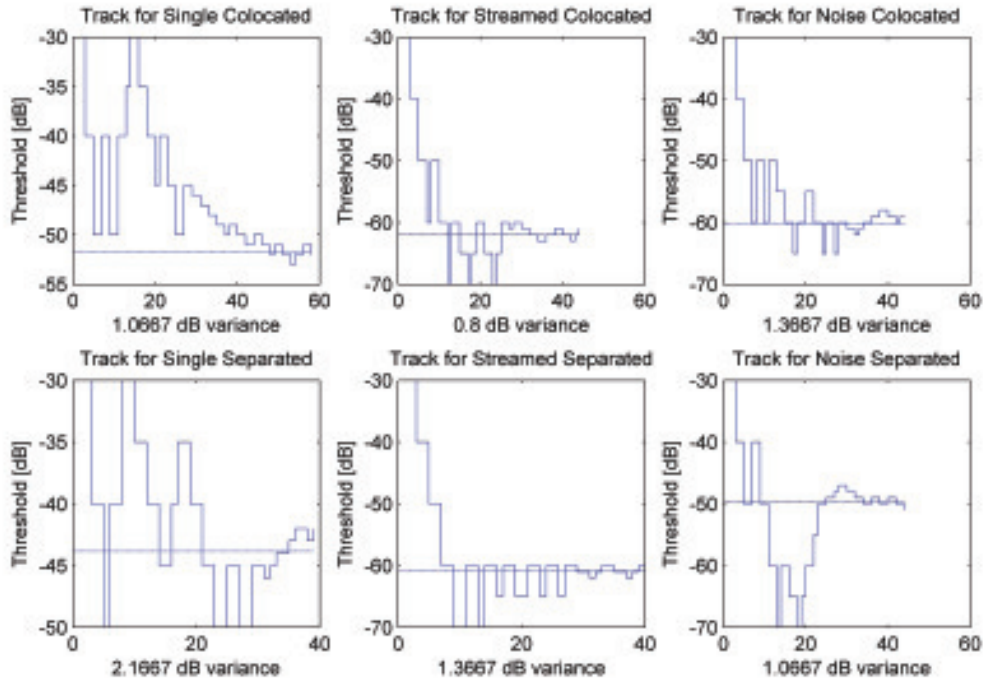


Figure A.26.: Tracks of the first session of subject 6. For details please refer to figure A.1.

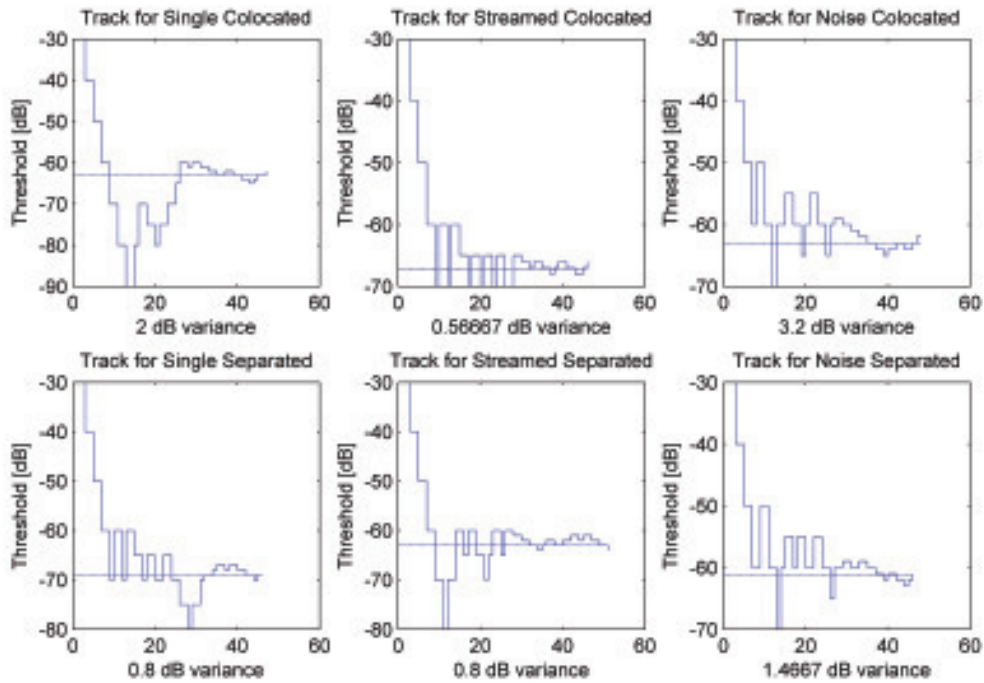


Figure A.27.: Tracks of the second session of subject 6. For details please refer to figure A.1.

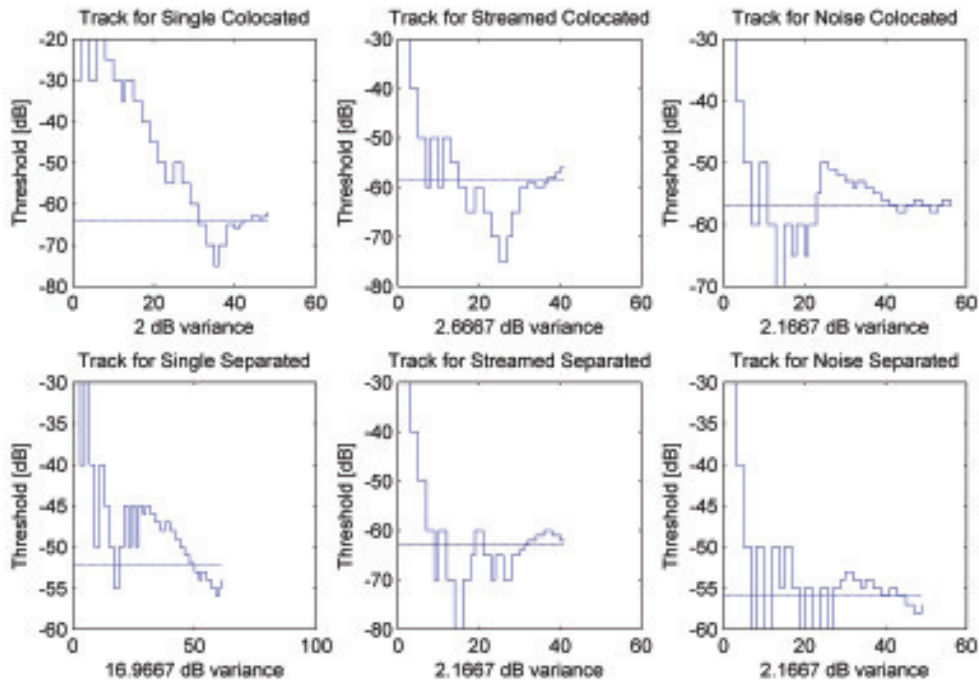


Figure A.28.: Tracks of the third session of subject 6. For details please refer to figure A.1.

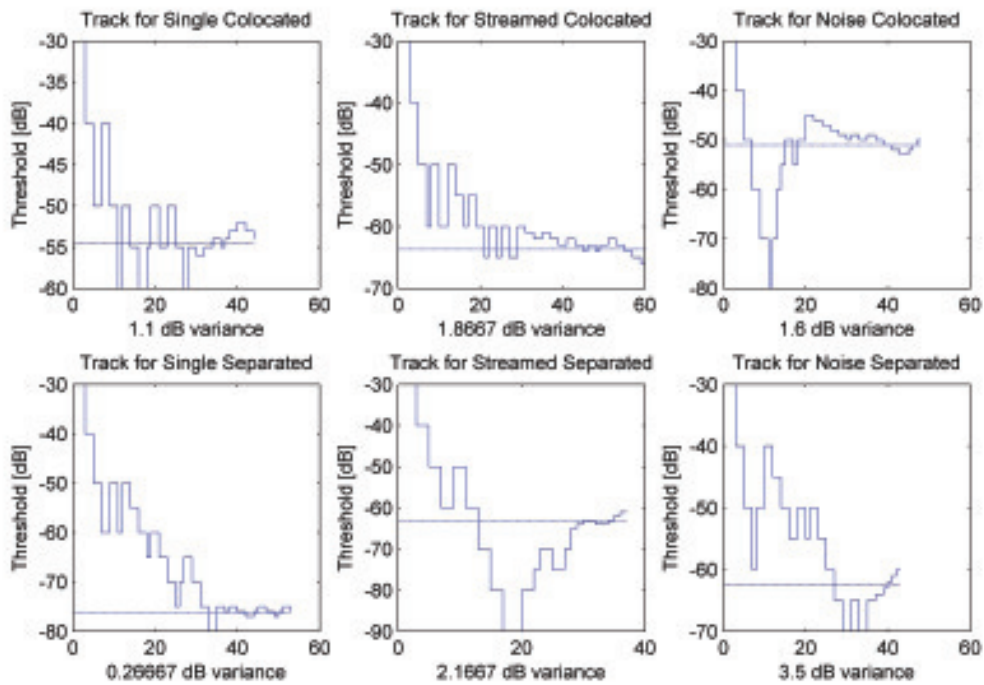


Figure A.29.: Tracks of the fourth session of subject 6. For details please refer to figure A.1.

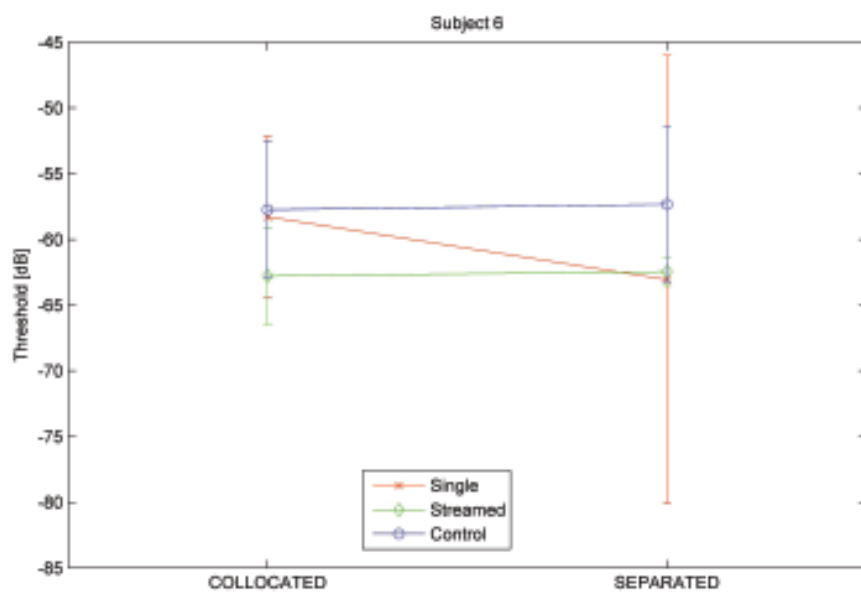


Figure A.30.: Results of subject 6. Please refer to figure A.6 for a description.

B. Individual figures of experiment 2

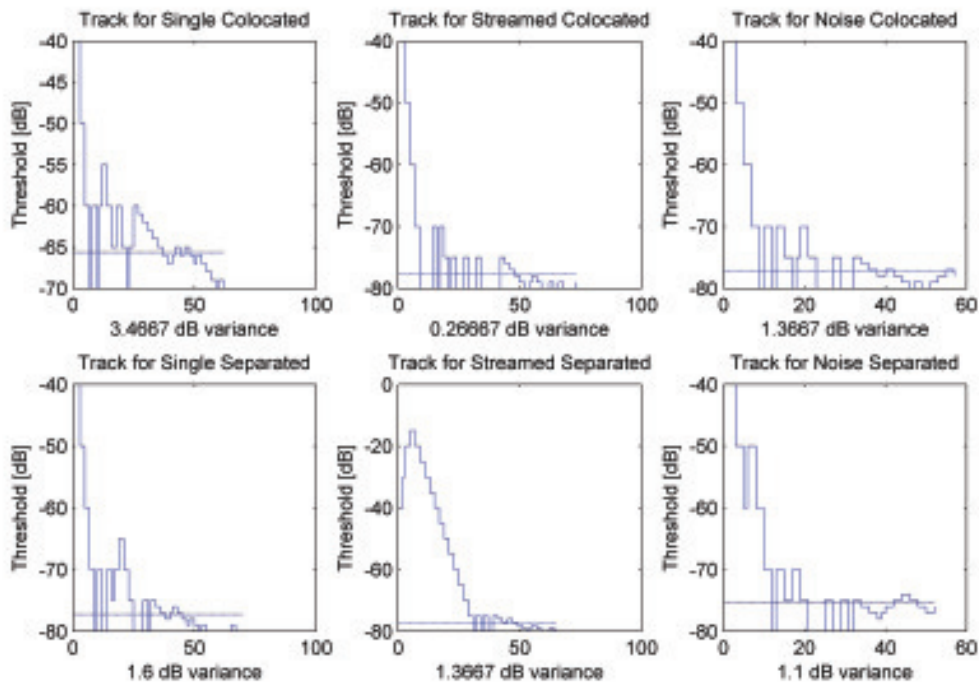


Figure B.1.: Tracks of the first session of subject 7. For details please refer to figure A.1.

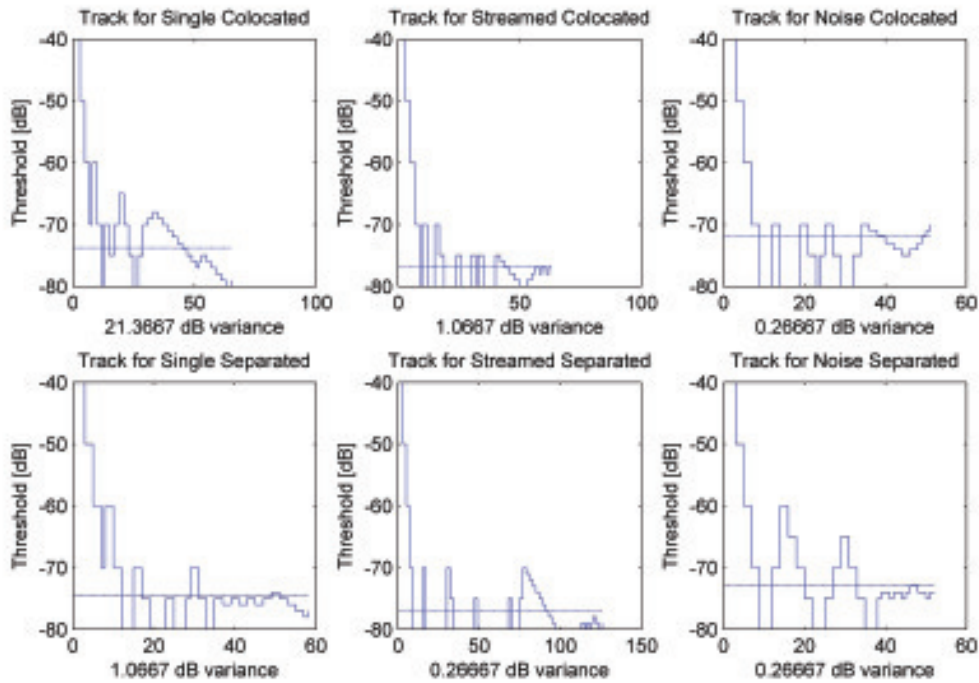


Figure B.2.: Tracks of the second session of subject 7. For details please refer to figure A.1.

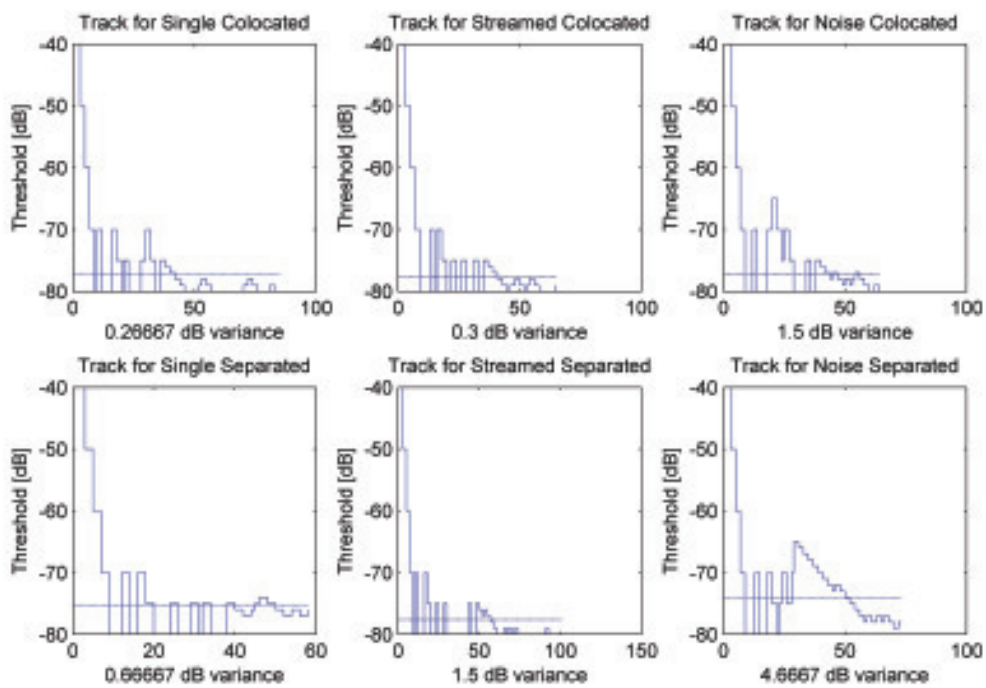


Figure B.3.: Tracks of the third session of subject 7. For details please refer to figure A.1.

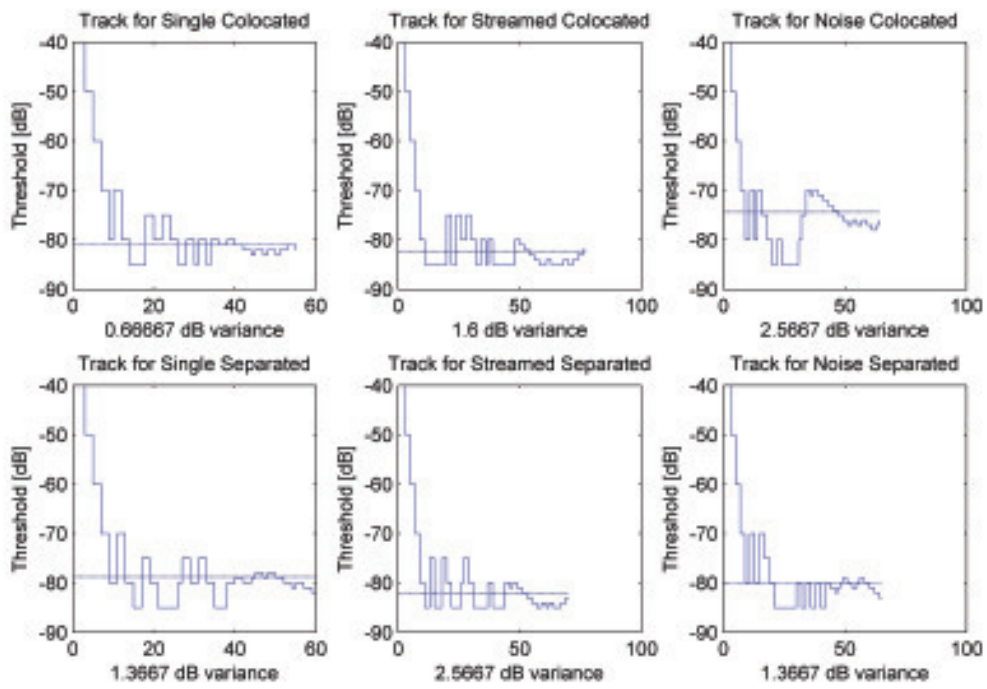


Figure B.4.: Tracks of the fourth session of subject 7. For details please refer to figure A.1.

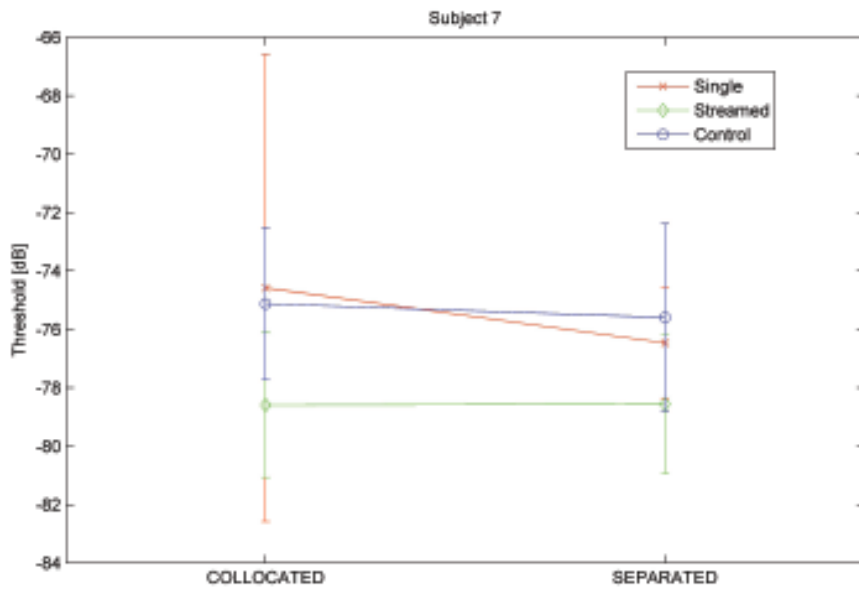


Figure B.5.: Results of subject 7. Please refer to figure A.6 for a description.

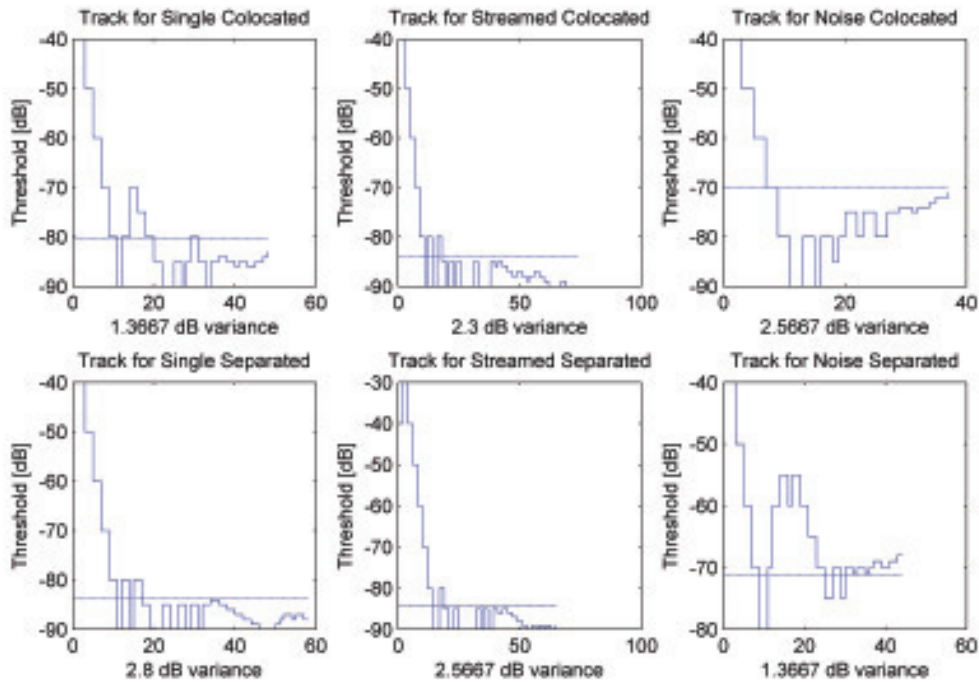


Figure B.6.: Tracks of the first session of subject 8. For details please refer to figure A.1.

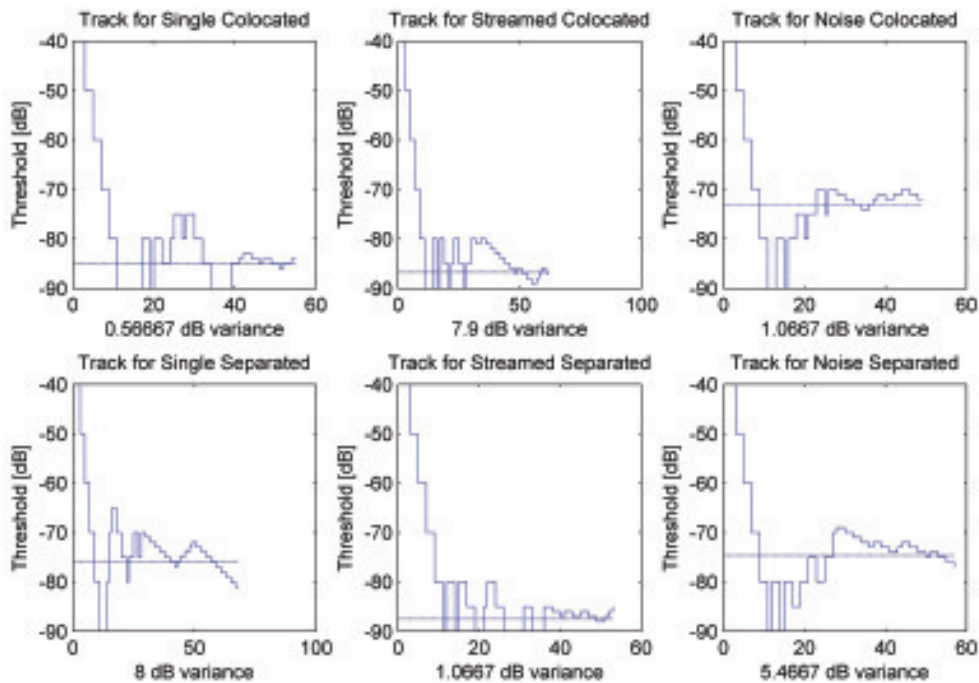


Figure B.7.: Tracks of the second session of subject 8. For details please refer to figure A.1.

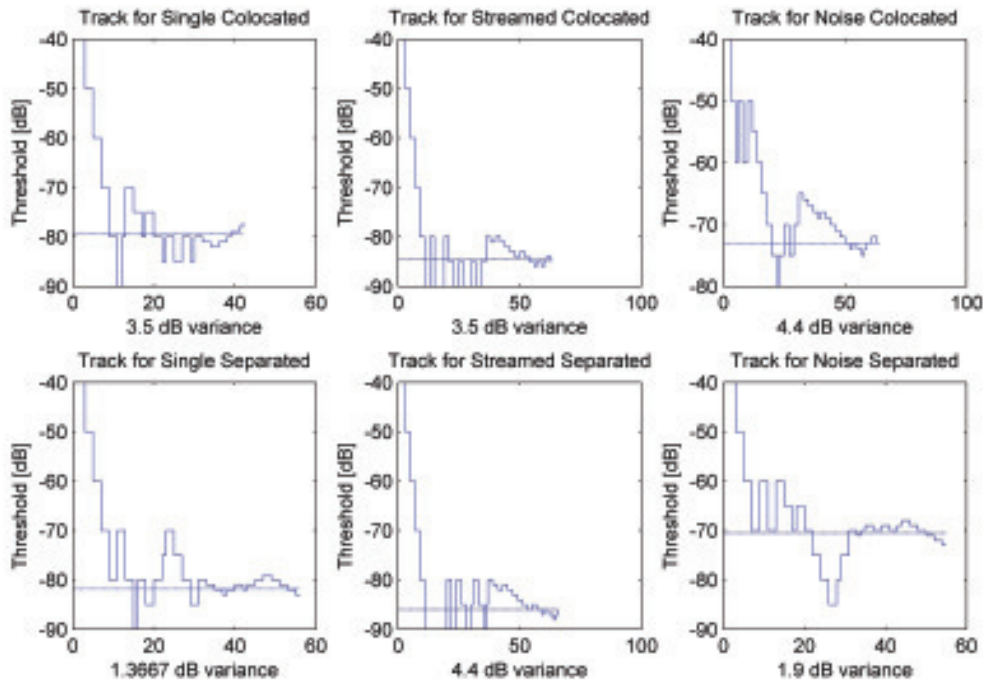


Figure B.8.: Tracks of the third session of subject 8. For details please refer to figure A.1.

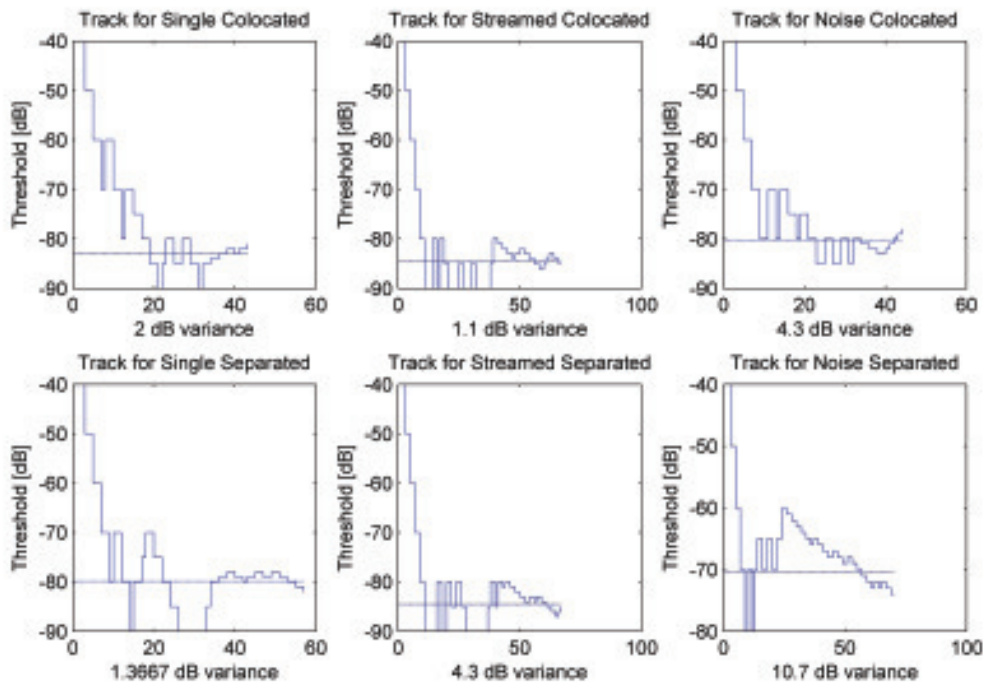


Figure B.9.: Tracks of the fourth session of subject 8. For details please refer to figure A.1.

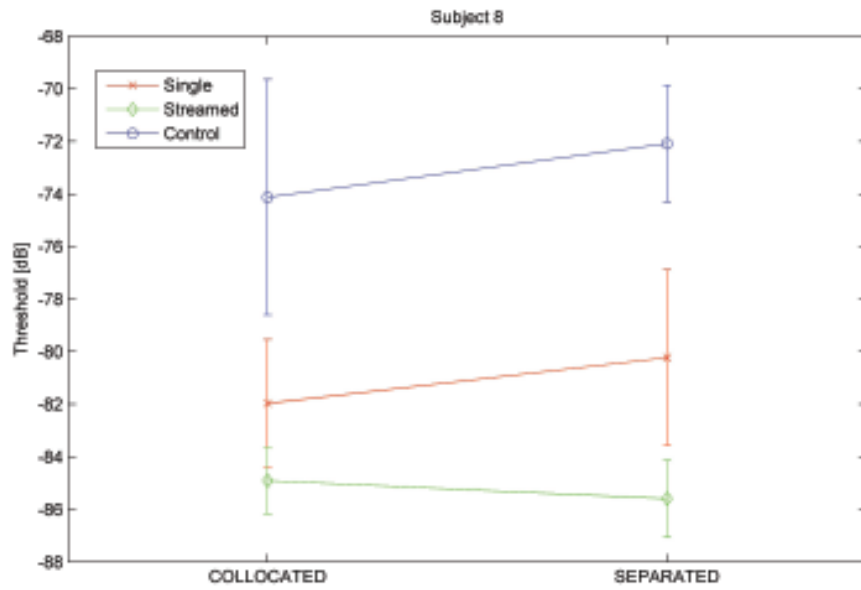


Figure B.10.: Results of subject 8. Please refer to figure A.6 for a description.

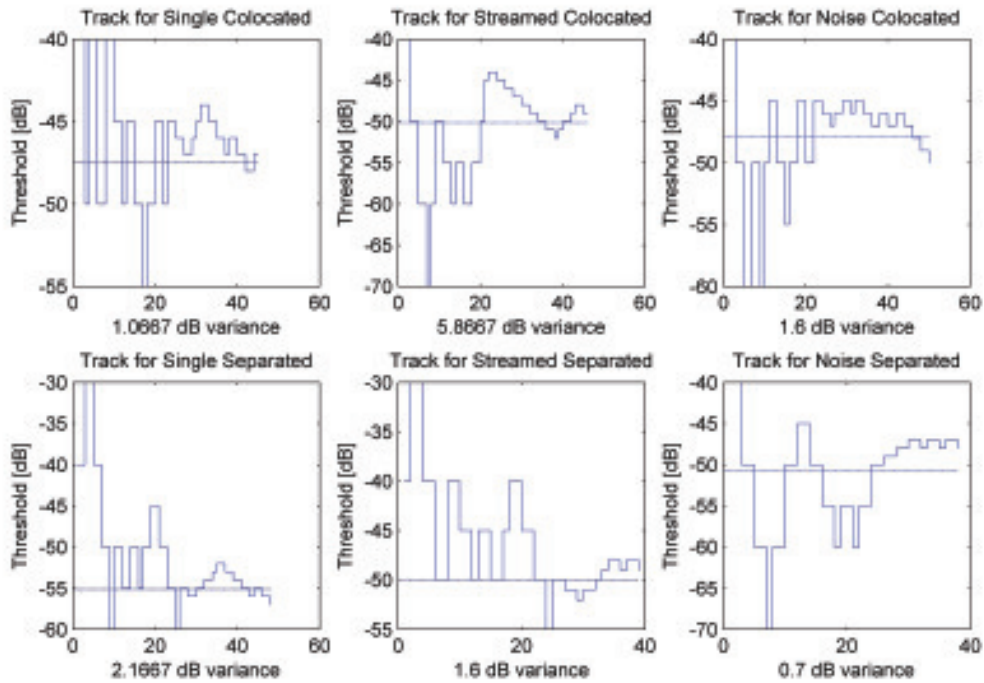


Figure B.11.: Tracks of the first session of subject 9. For details please refer to figure A.1.

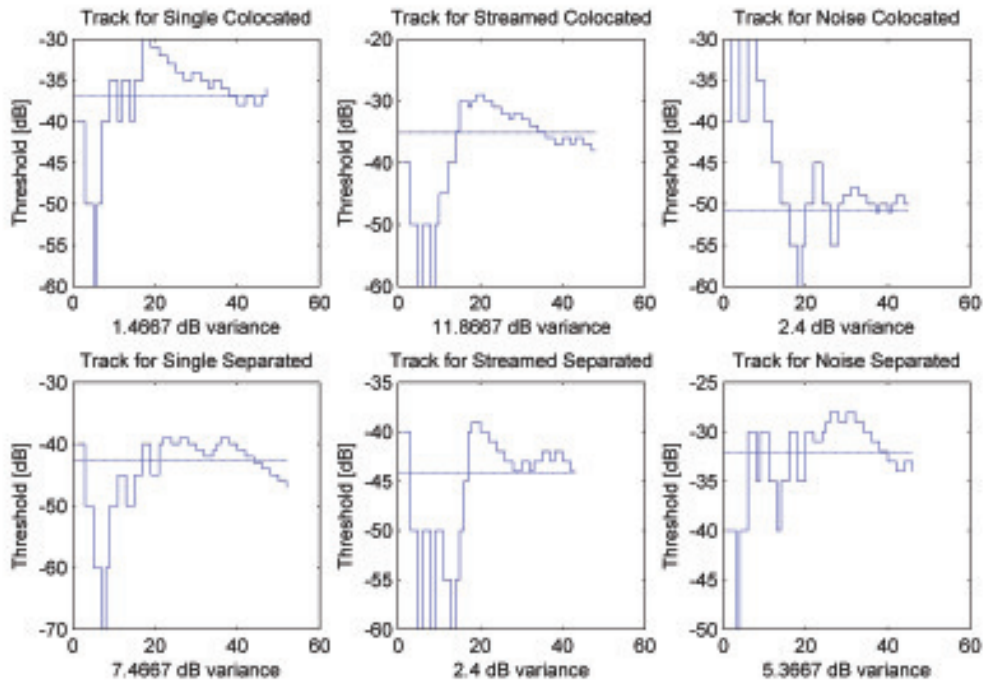


Figure B.12.: Tracks of the second session of subject 9. For details please refer to figure A.1.

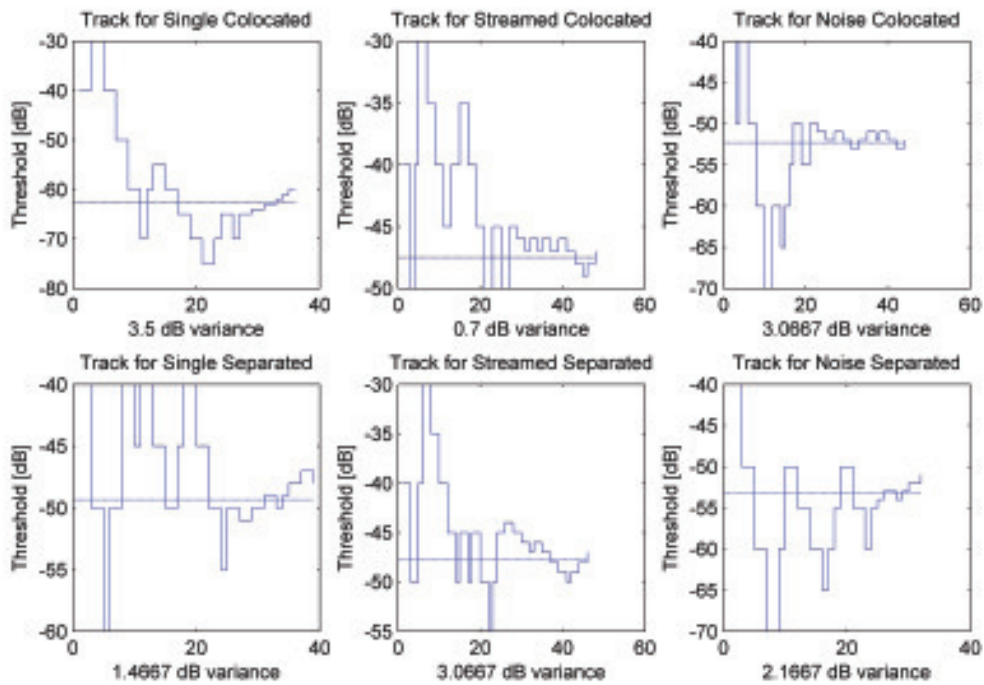


Figure B.13.: Tracks of the third session of subject 9. For details please refer to figure A.1.

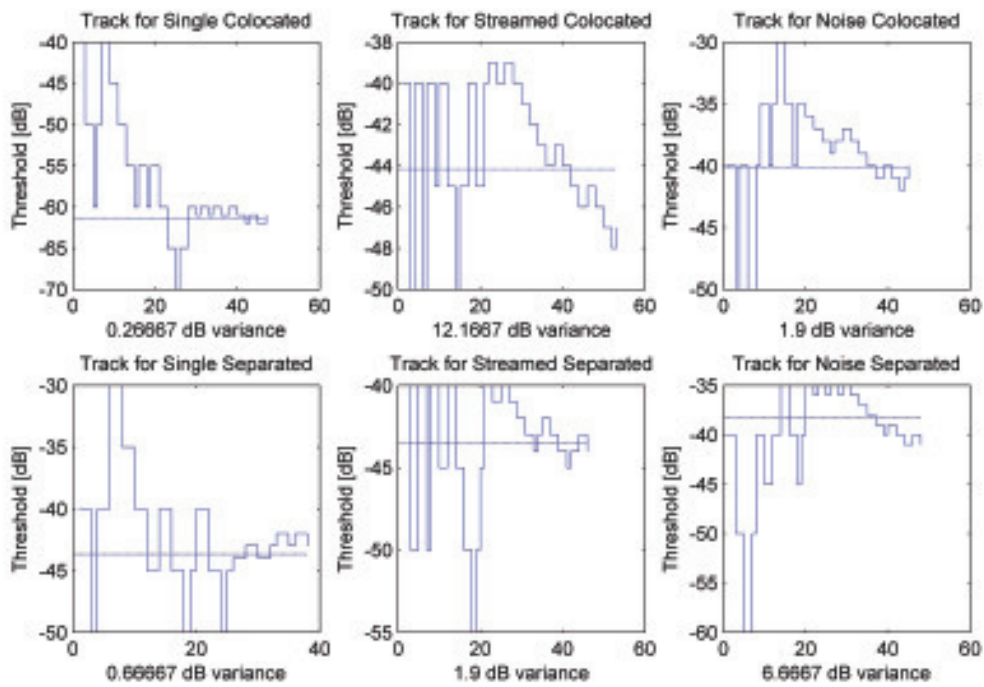


Figure B.14.: Tracks of the fourth session of subject 9. For details please refer to figure A.1.

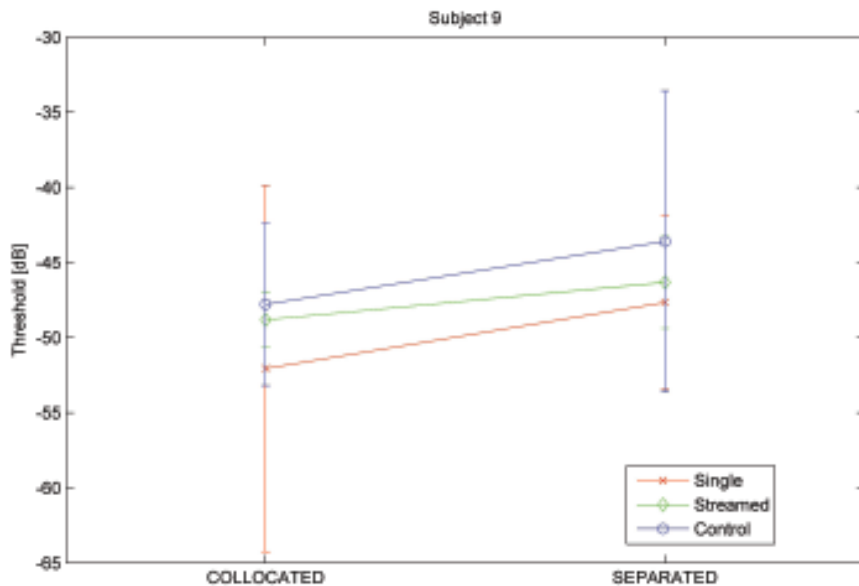


Figure B.15.: Results of subject 9. Please refer to figure A.6 for a description.

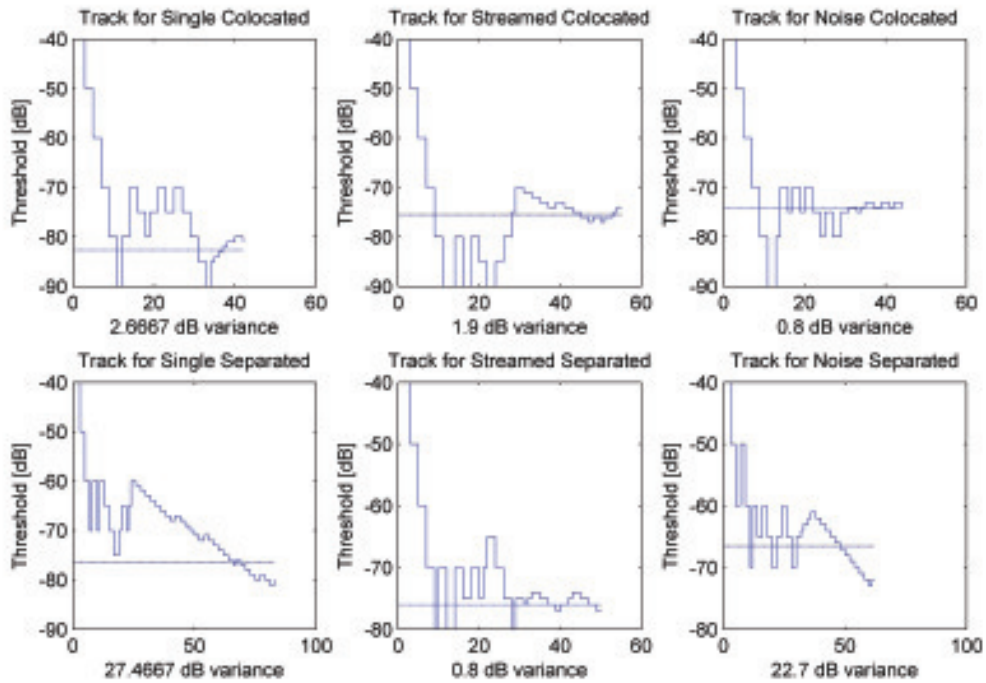


Figure B.16.: Tracks of the first session of subject 10. For details please refer to figure A.1.

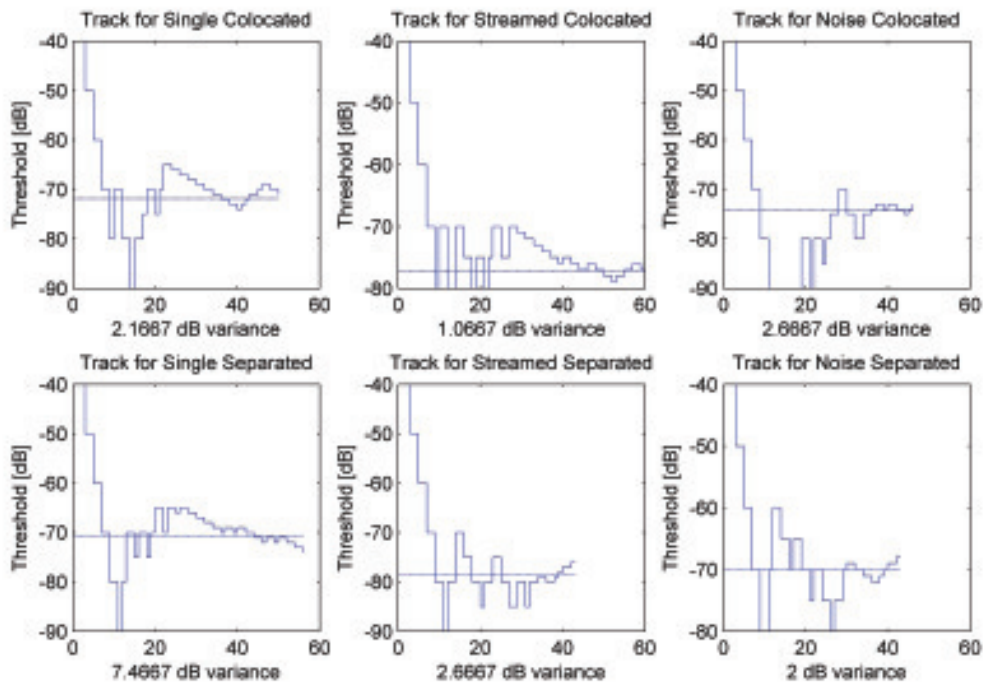


Figure B.17.: Tracks of the second session of subject 10. For details please refer to figure A.1.

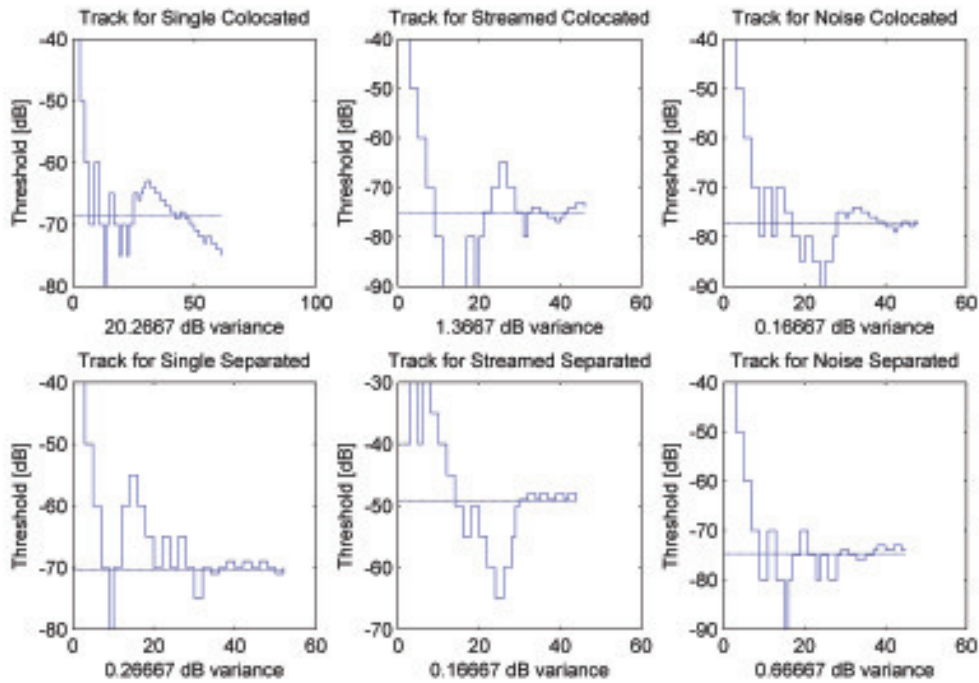


Figure B.18.: Tracks of the third session of subject 10. For details please refer to figure A.1.

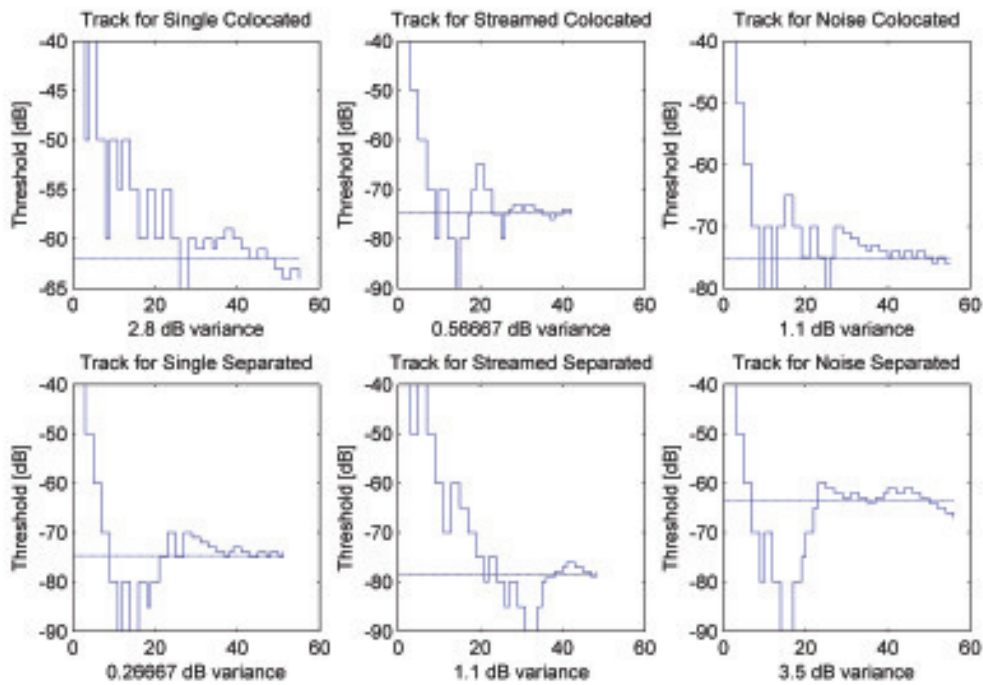


Figure B.19.: Tracks of the fourth session of subject 10. For details please refer to figure A.1.

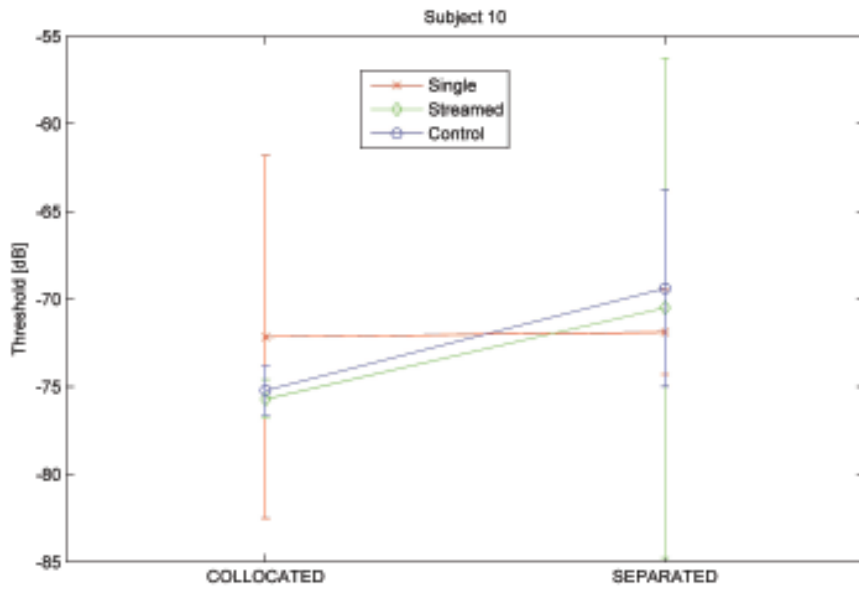


Figure B.20.: Results of subject 10. Please refer to figure A.6 for a description.

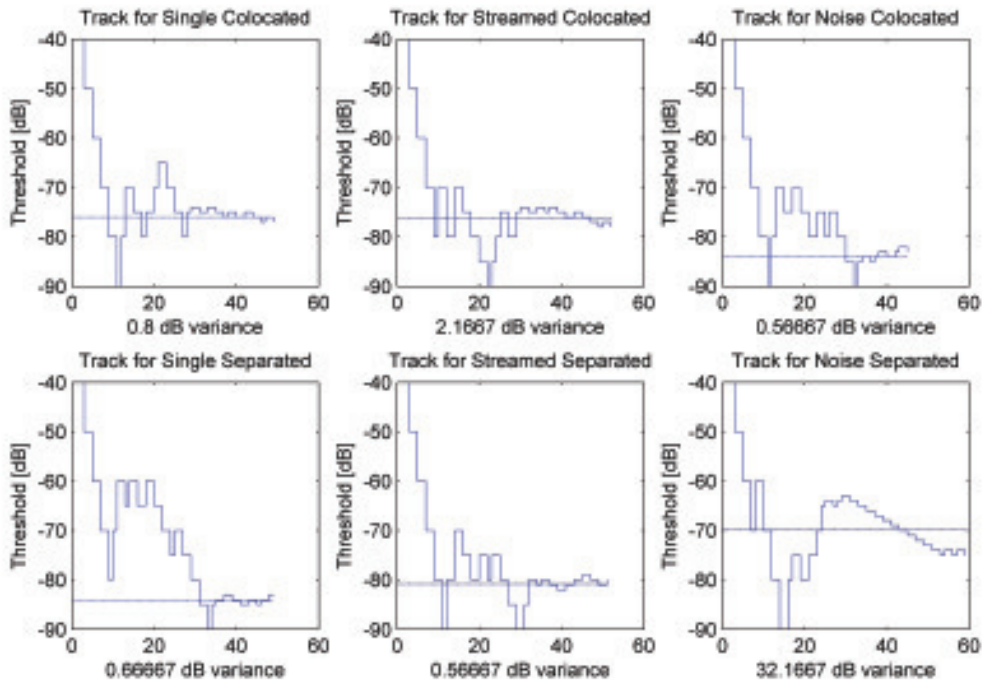


Figure B.21.: Tracks of the first session of subject 11. For details please refer to figure A.1.

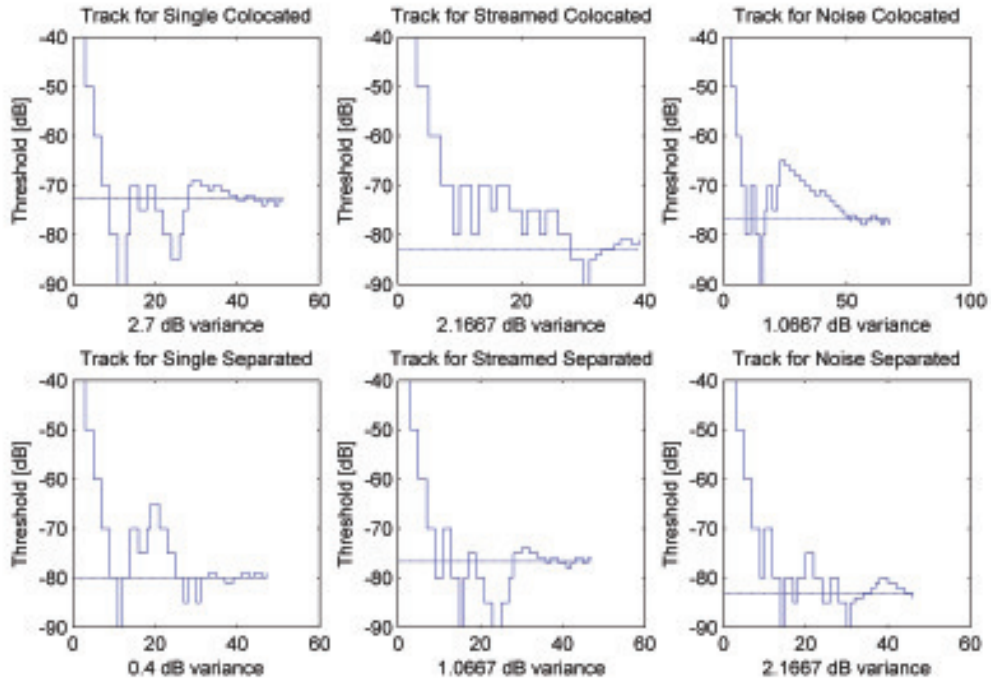


Figure B.22.: Tracks of the second session of subject 11. For details please refer to figure A.1.

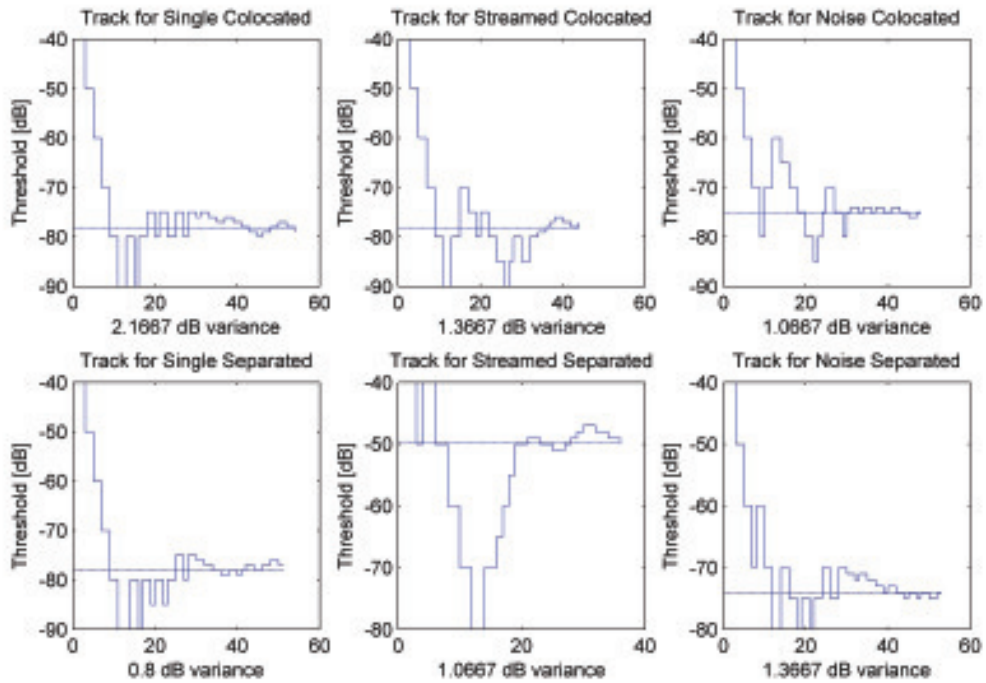


Figure B.23.: Tracks of the third session of subject 11. For details please refer to figure A.1.

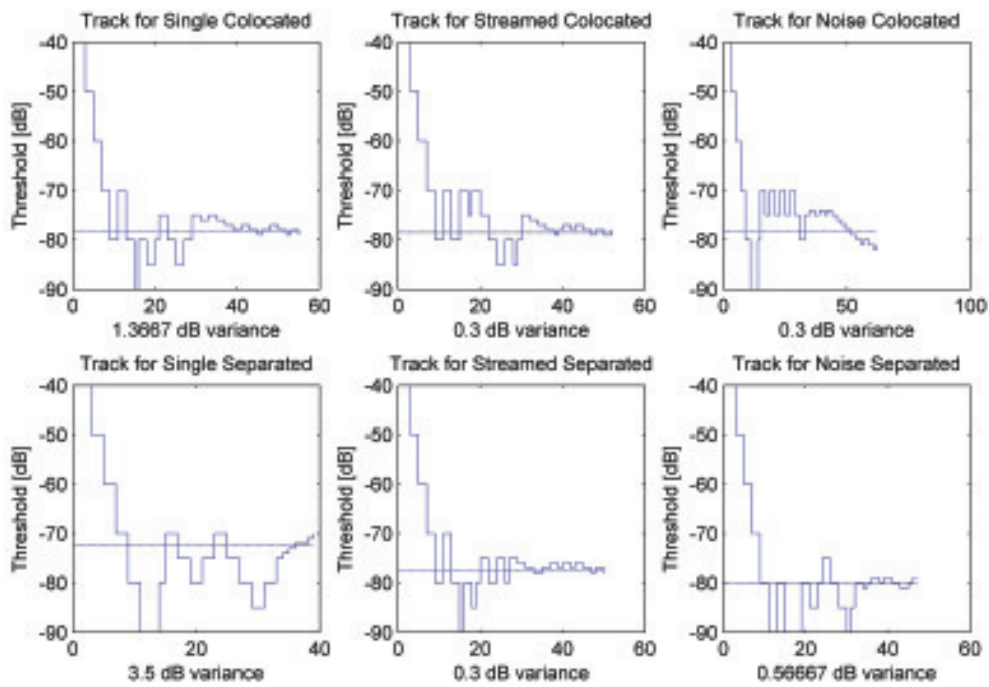


Figure B.24.: Tracks of the fourth session of subject 11. For details please refer to figure A.1.

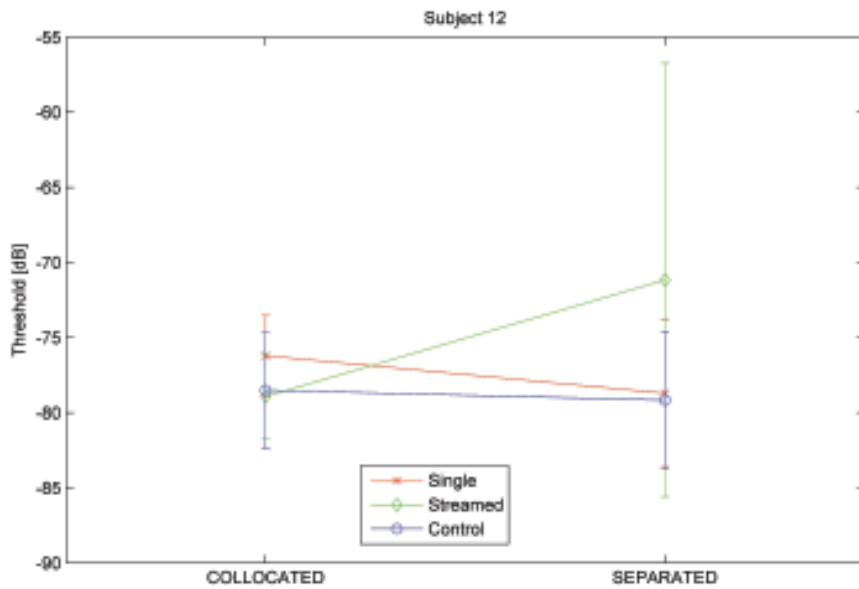


Figure B.25.: Results of subject 11. Please refer to figure A.6 for a description.

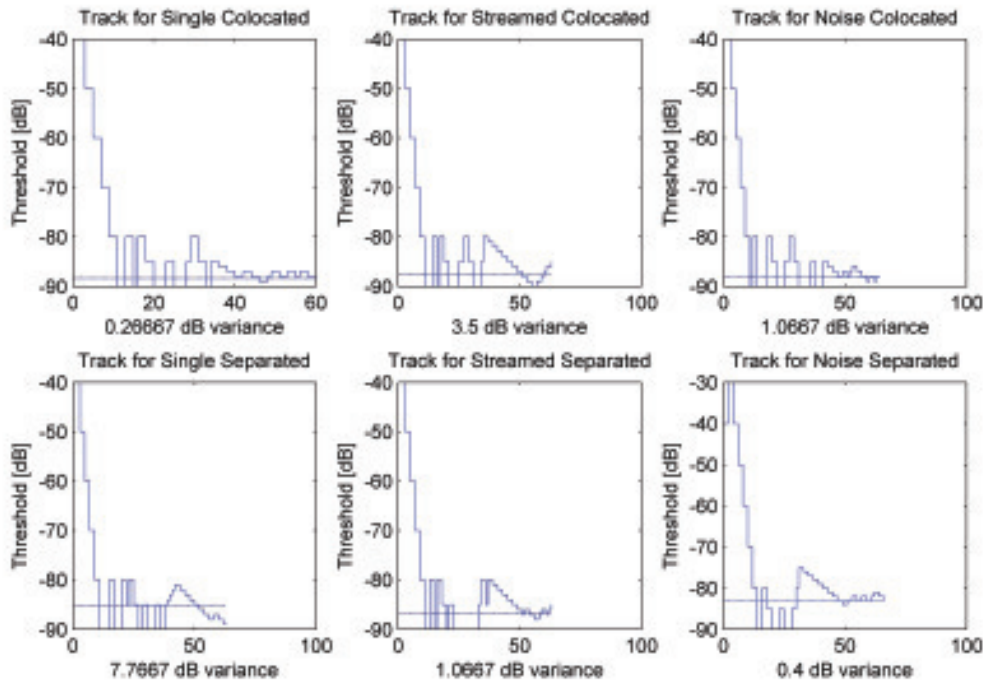


Figure B.26.: Tracks of the first session of subject 12. For details please refer to figure A.1.

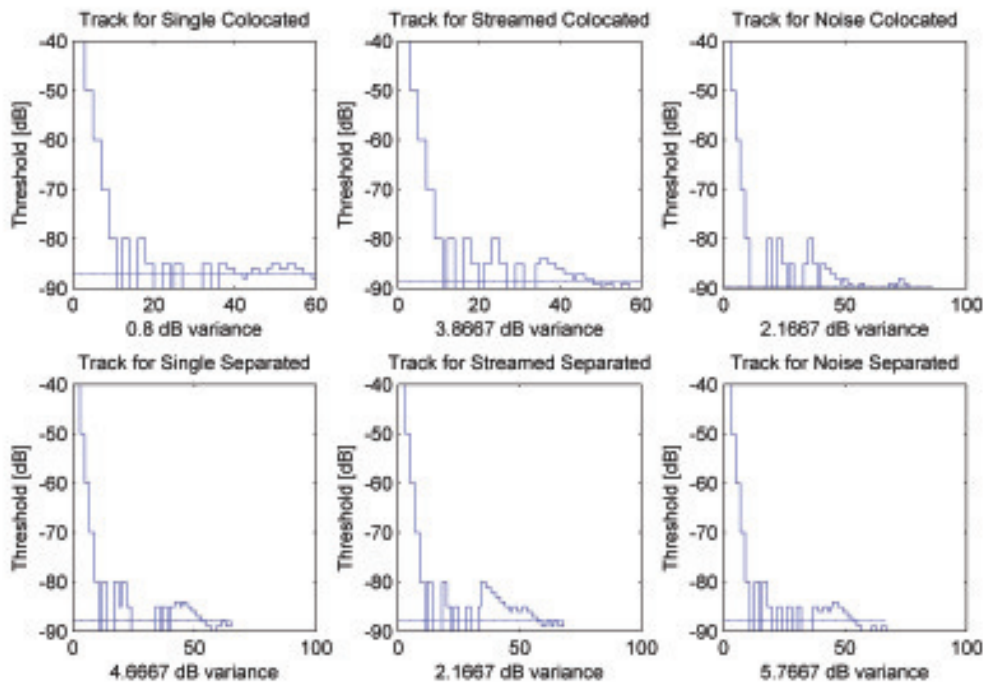


Figure B.27.: Tracks of the second session of subject 12. For details please refer to figure A.1.

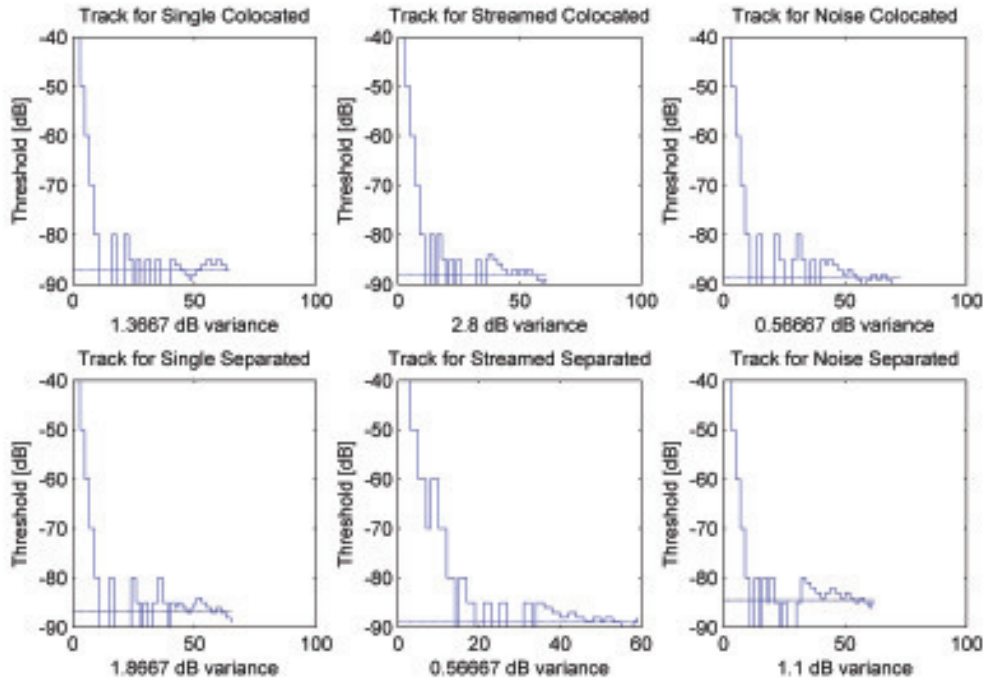


Figure B.28.: Tracks of the third session of subject 12. For details please refer to figure A.1.

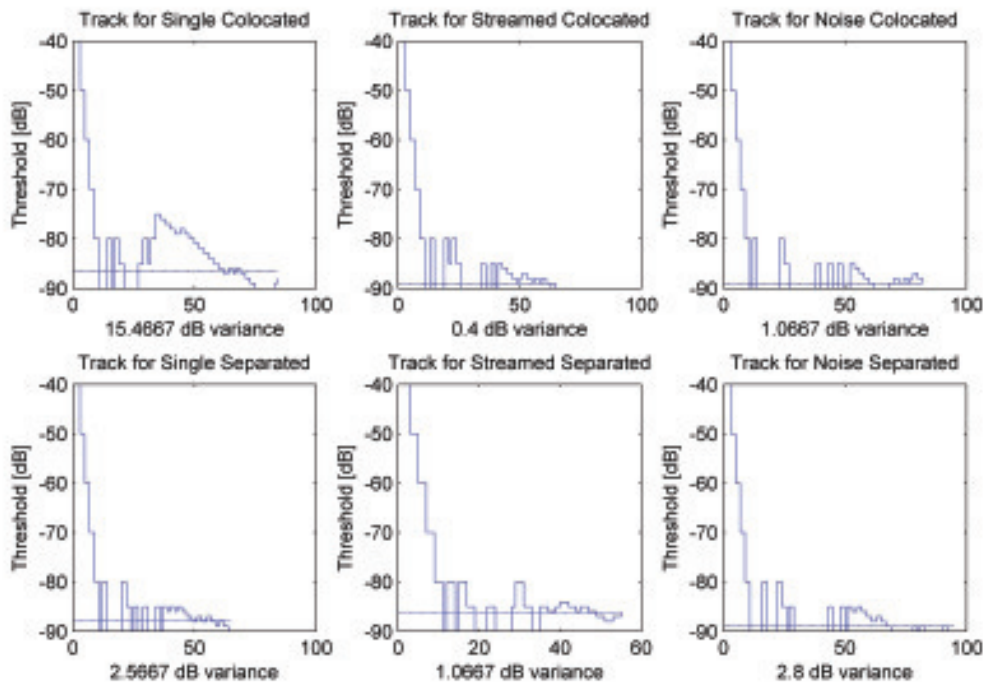


Figure B.29.: Tracks of the fourth session of subject 12. For details please refer to figure A.1.

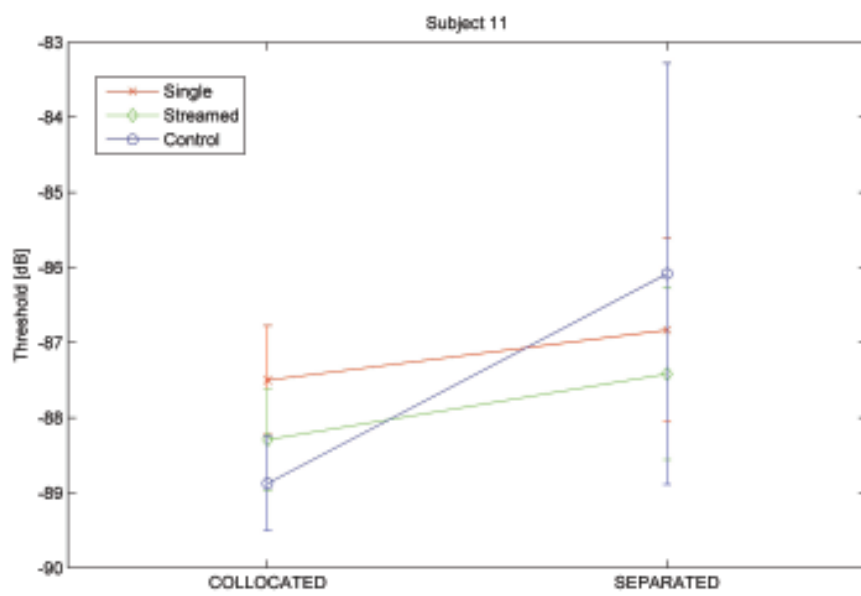


Figure B.30.: Results of subject 12. Please refer to figure A.6 for a description.

Bibliography

- [1] O. Santala, V. Pulkki
Resolution of spatial distribution perception with distributed sound source in anechoic conditions
Convention Paper 7779, 126. AES Convention, Munich, 2009.
- [2] E. Zwicker, H. Fastl
Psychoacoustics: Facts and Models
Springer Verlag, Berlin, 1999.
- [3] B.G. Shinn-Cunningham
Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays
Proceedings of the 2002 International Conference on Auditory Display, Kyoto, 2002.
- [4] N. Kopçô, B.G. Shinn-Cunningham
Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment
JASA 115 (5), November 2003.
- [5] V. Best
Spatial Hearing with Simultaneous Sound Sources: A Psychophysical Investigation
PhD Thesis at The University of Sidney, April 2004.
- [6] A. Sontacchi, R. Höldrich
Distance Coding in 3D Sound Fields
21. AES Conference, St. Petersburg, 2002.
- [7] W.R. Thurlow, J.W. Mangels, P.S. Runge
Head Movements during sound localization
Journal of the ASA Issue 42, Seiten 489-493, 1967.
- [8] W. Schirmer
Die Richtcharakteristik des Ohres
Hochfrequenztechnik und Elektroakustik Vol. 72, Seite 39, 1963.

- [9] G. Jahn
Zum Unterschied zwischen einohrigem und zweiohrigem Hören
Hochfrequenztechnik und Elektroakustik Vol. 67-3, Seite 69, 1958.
- [10] W. Ahnert, W. Reichardt
Grundlagen der Beschallungstechnik
S. Hirzel Verlag, Stuttgart, 1981.
- [11] J. Blauert
Räumliches Hören
S. Hirzel Verlag, Stuttgart, 1976.
- [12] J. Blauert (Ed.)
Communication Acoustics
Springer Verlag, Berlin, 2005.
- [13] J. M. Buchholz, J. Mourjopoulos, J. Blauert
Room Masking: Understanding and Modelling the Masking of Room Reflections
Convention Paper 5312, 110. AES Convention, Amsterdam, 2001.
- [14] J. M. Buchholz
Characterizing the monaural and binaural processes underlying reflection masking
Hearing Research Issue 232, Seiten 52-66, 2007.
- [15] T. Dau, D. Püschel, A. Kohlrausch
A quantitative model of the "effective" signal processing in the auditory system
Journal of the ASA Issue 99, Seiten 3615-3622, 1997.
- [16] M. Takanen et al.
Localization of Consecutive Sound Events in Reverberant Environment
Convention Paper 7807, 126. AES Convention, Munich, 2009.
- [17] W. Burgdorf
Untersuchungen zur Wahrnehmbarkeit verzögerter Schallsignale
Acoustica Vol. 11, Seiten 97-111, 1961.