Supplemental Materials

Here we report the results obtained in each Experiment separately, and then move on to provide additional analysis on the full dataset. Only the features that changed between Experiments are detailed below. A full description of the methods can be found in the main manuscript. The scales of all graphs were kept the same as the main manuscript where to ease comparisons.

Table 4. Mean thresholds and standard deviations for the 10 assessments utilized plus the spatial release metric in PART's native measurement units for Experiment 1.

Test	*Exp1 M (SD)*	*Exp2 M (SD)*	*Exp3 M (SD)*	Units
Gap	2.16 (3.12)	2.31 (3.44)	3.18 (3.51)	Gap length (ms)
Dichotic FM	0.62 (2.35)	0.51 (2.28)	0.48 (2.6)	Modulation Depth (hz)
Diotic FM	7.08 (1.69)	6.38 (2.04)	5.77 (1.92)	Modulation Depth (hz)
TM	1.64 (1)	1.57 (.83)	1.56 (1.41)	Modulation depth (dB)
SM	1.5 (.76)	1.66 (1.12)	1.96 (1.32)	Modulation depth (dB)
STM	1.05 (.73)	1.12 (.91)	1.39 (1.38)	Modulation depth (dB)
No-Notch	-11.51 (1.57)	-12.35 (2.22)	-11.08 (3.44)	Signal-to-masker ratio (dB)
Notch	-31.66 (2.67)	-32.49 (4.07)	-29.65 (9.7)	Signal-to-masker ratio (dB)
SR Co-located	2.17 (1.44)	1.95 (1.83)	0.34 (4.19)	Signal-to-masker ratio (dB)
SR Separated	-4.74 (3.24)	-4.54 (2.96)	-3.69 (4.26)	Signal-to-masker ratio (dB)
Spatial Release	6.92 (3.44)	6.49 (2.94)	4.04 (3.71)	SR (Sep - Co) (dB)

Experiment 1 (standard; n=51)

In this experiment a 2:1 step size ratio between the step-up and the step-down behaviors of the staircase was used. All participants in this wave were tested in silent conditions using Sennheiser 280 Pro circumaural headphones.

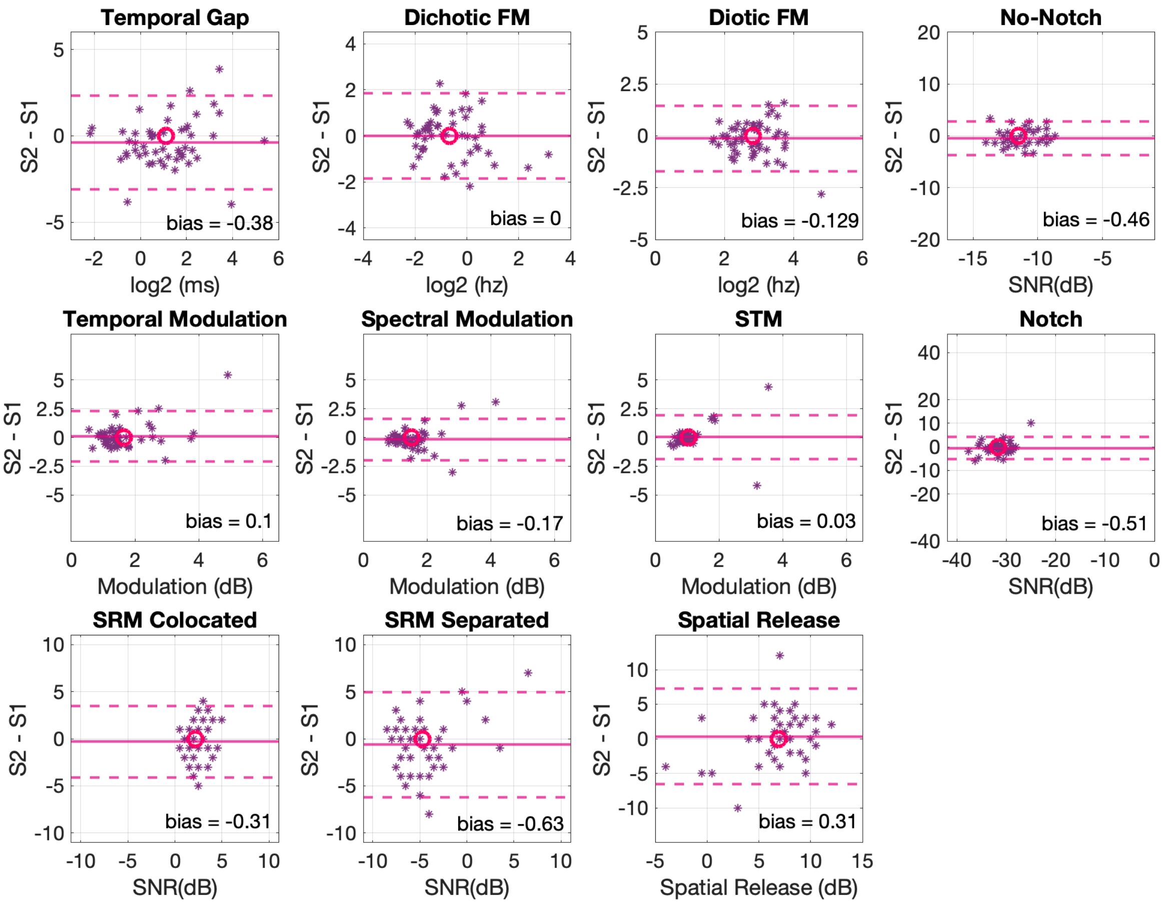


Figure 10. Limits of agreement of the estimated thresholds between sessions for all tests in Experiment 1. The solid lines indicate the mean difference between sessions. Dotted lines indicate the 95% limits of agreement. The red circle indicates the mean threshold for each test centered at cero difference between sessions. Solid lines below cero indicate better performance on session 2 (except the spatial release metric).

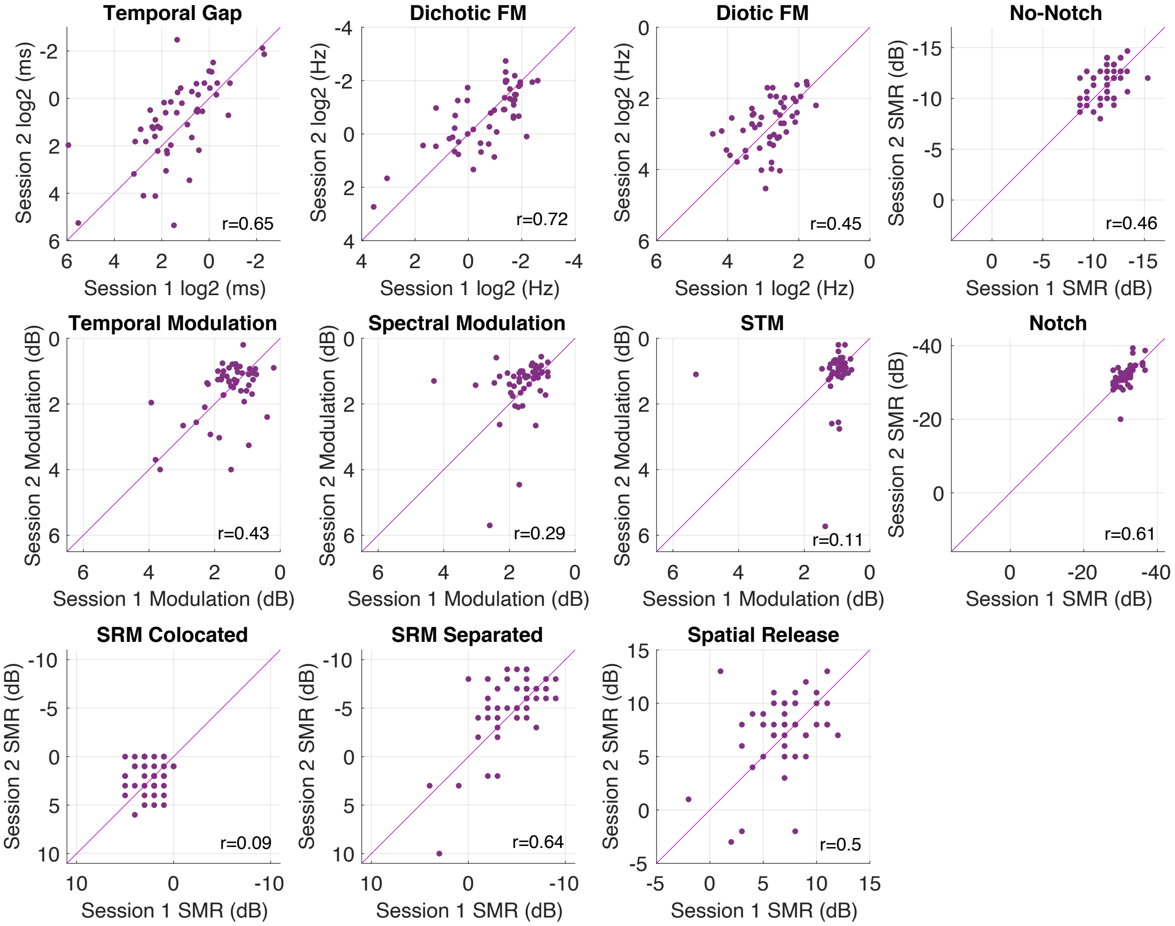


Figure 11. Scatter plots of Session 1 vs Session 2 for the 10 assessments in Experiment 1. All axes are oriented to show better performance values away from the origin. Correlations are indicated in the lower right of each panel. The diagonal is plotted to ease evaluation of differences between sessions, dots above this line indicate better performance in session 2.

The main differences that can be observed between these threshold estimates results and those reported in the main manuscript come from the distribution of the outliers. No extreme outliers were present in the case of the No-Notch test and in consequence the limits of agreement are reduced (see Table 5). The non-significant correlations for the Spatial Release Co-located condition and the STM can be explained by the reduced between-subjects variability rather than reduced reliability and support the use of the limits of agreement as the main analysis as suggested by Altman & Bland (1983).

Table 5. Significance testing for all assessments in Experiment 1 at two time points. * indicate significance at α = .05

Test	*Bias*	*Limits of Agreement*	Units	*r (p)*	*t (p)*	*df*	*Cohen's d*
Gap	-0.37	[-3.08, 2.33]	Log2 (ms)	.65 (<.01)*	1.94 (.058)	50	0.27
DichoticFM	-0.001	[-1.85, 1.85]	Log2 (hz)	.72 (<.01)*	0.002 (.99)	50	<0.001
DioticFM	-0.12	[-2, 1.81]	Log2 (hz)	.45 (<.01)*	1.14 (.25)	50	0.16
TM	0.1	[-2.06, 2.27]	M (dB)	.43(<.01)*	-0.63 (.53)	50	-0.08
SM	-0.16	[-1.97, 1.63]	M (dB)	.29 (.04)*	1.38 (.17)	50	0.19
STM	0.02	[-1.89, 1.95]	M (dB)	.11 (.42)	-0.2 (.83)	50	-0.02
No-Notch	-0.45	[-3.6, 2.75]	SMR (dB)	.46 (<.01)*	1.99 (.052)	50	0.27
Notch	-0.5	[-5.3, 4.28]	SMR (dB)	.61 (<.01)*	1.48 (.14)	50	0.2
Co-located	-0.31	[-4.12, 3.49]	SMR (dB)	.09 (.51)	1.15 (.25)	50	0.16
Separated	-0.62	[-6.19, 4.94]	SMR (dB)	.64 (<.01)*	1.57 (.12)	50	0.22
SpatialR	0.31	[-6.59, 7.22]	SMR (dB)	.5 (<.01)*	-0.63 (.52)	50	-0.08

Experiment 2 (silence; n=51)

In this experiment a 1.5:1 step size ratio between the step-up and the step-down behaviors of the staircases of adaptive parameters was used to increase the sensitivity of our measures. This was not the case of the spatial release tasks, whose progressive structure remained unchanged. Furthermore, half of our participants were tested using Sennheiser 285 Pro circumaural headphones on session 1 and the other half used Bose noise cancelling circumaural headphones. This was counterbalanced across sessions. Our aim with this manipulation was to see if our results on experiment 1 generalized to a headset with active noise cancelling. We kept this counterbalanced design to both replicate Experiment 1 (perhaps with more sensitivity) and measure the effects of the type of headphones used. The rest of the experimental settings remained unchanged from those of Experiment 1.

Outlier rejection

As reported in the main manuscript, we rejected the performance of one participant in the Notch task. Figure 12 shows the only analysis conducted on this data to the sole purpose of justifying its rejection. The values for the adaptive parameter masker level are distributed on the y-axis and the trial numbers on the x-axis. Each line in the figure represents a single participant going through a single session. The black bold line represents the participant whose threshold estimates were rejected from further analysis. This participant had a performance similar to the rest of the participants on session 1, but in session 2 provides almost exclusively incorrect responses. This performance yields a threshold estimate beyond 8 SD away from the mean and does not relate to this person hearing ability in a meaningful way. Since the reliability of our measures across a range of conditions is the main aim of this work, we rejected the both sessions for this test as well as this participant's composite score from further analysis.

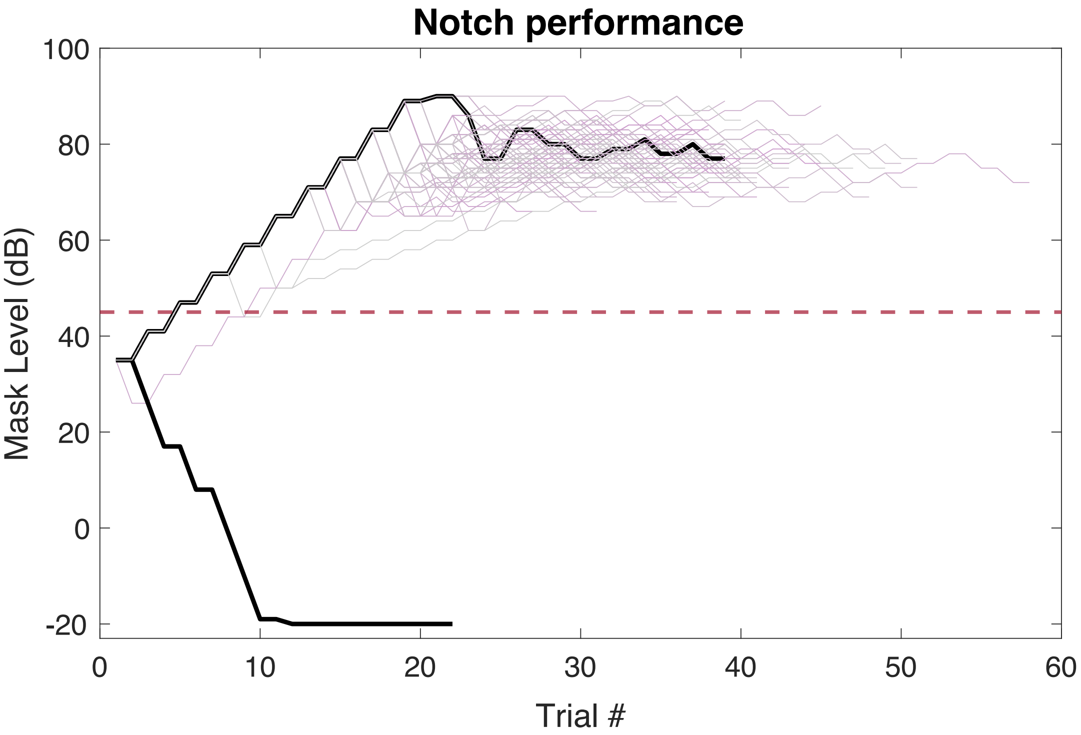


Figure 12. Shows adaptive track progressions for all participants performing the Notch task in two sessions. Bold black line corresponds to participant #49 rejected from further analysis. The dotted red line shows the level of the target tone at 2 kHz.

The main difference we find with these threshold estimates results and those reported in the main manuscript for the aggregated dataset come from the Spatial Release Task. The learning effect we reported in the main manuscript manifested strongest in this experiment and shows a .6 difference in terms of SD or almost 2 dB improvement. Other small but significant differences occurred in the diotic FM and the No-Notch tasks (see Table 6). Variations in the magnitude of the correlations do not seem meaningful in the face of the stability observed in the limits of agreement analysis and is again interpreted as arising from a between-subject variability that approximates the magnitude of measurement error.

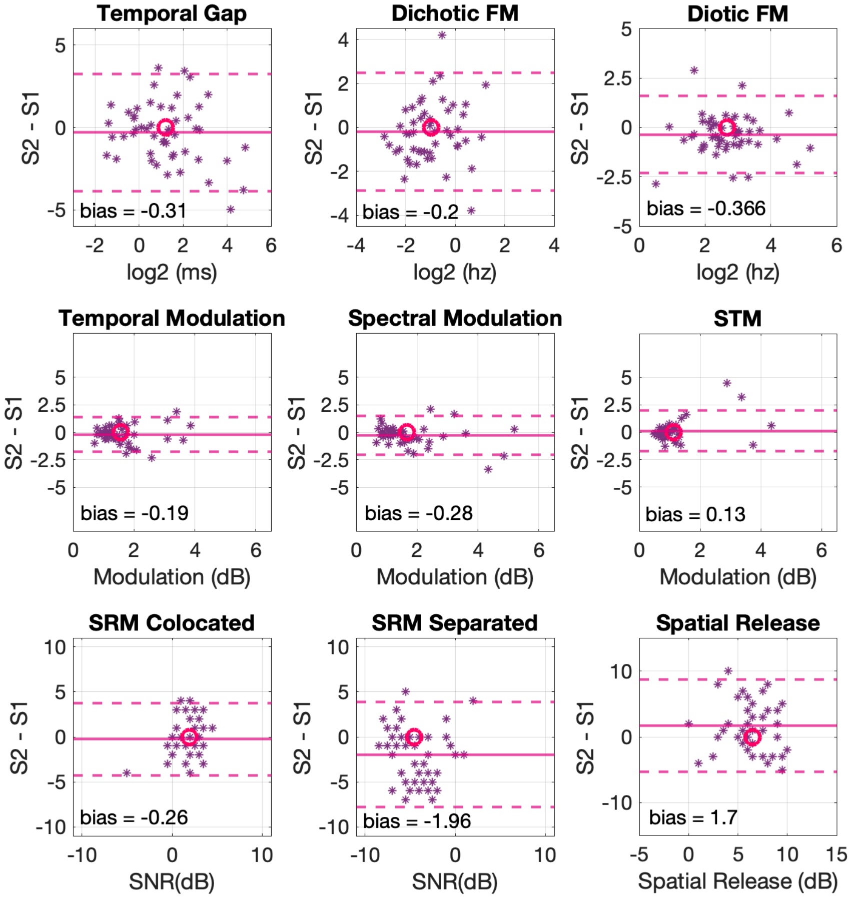


Figure 13. Limits of agreement of the estimated thresholds between sessions for all tests in Experiment 2. The solid lines indicate the mean difference between sessions. Dotted lines indicate the 95% limits of agreement. The red circle indicates the mean threshold for each test centered at cero difference between sessions. Solid lines below cero indicate better performance on session 2 (except the spatial release metric).

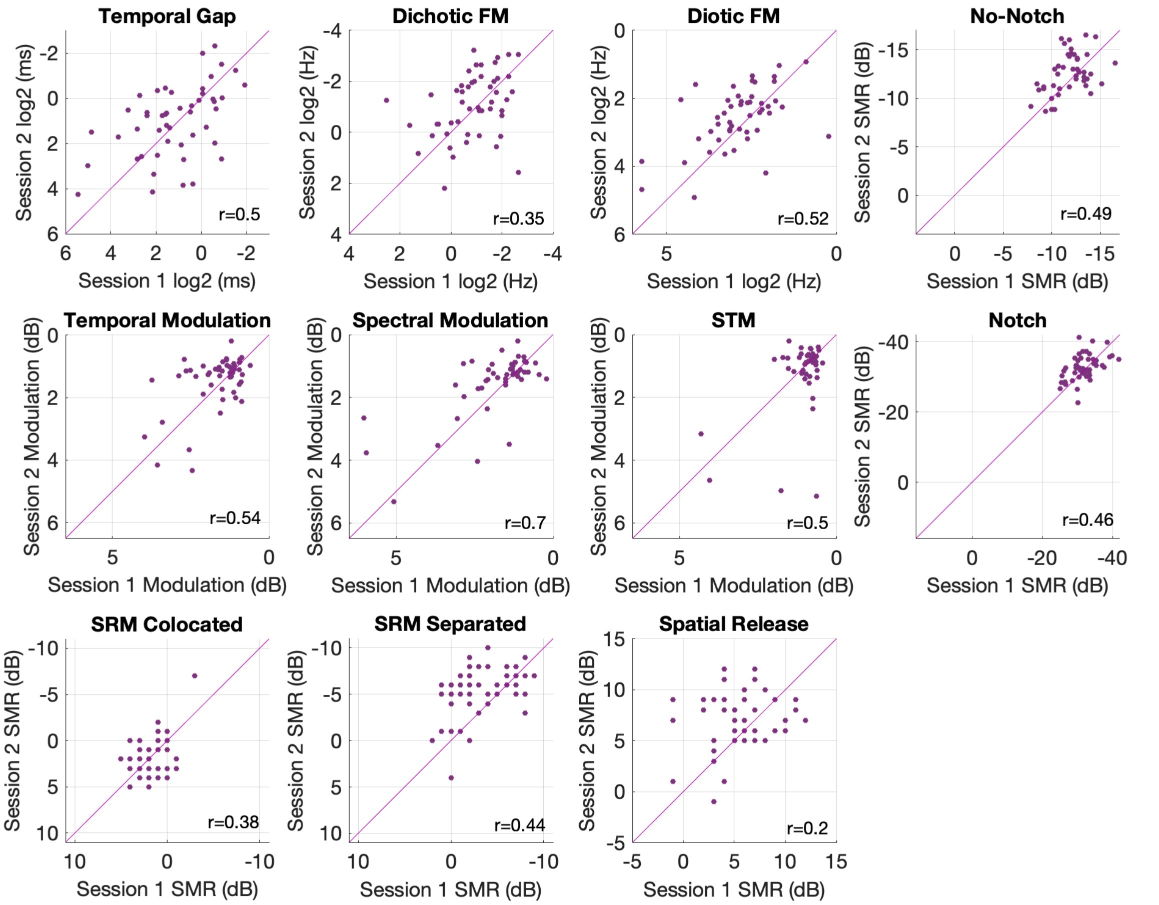


Figure 14. Scatter plots of Session 1 vs Session 2 for the 10 assessments in Experiment 2. All axes are oriented to show better performance values away from the origin. Correlations are indicated in the lower right of each panel. The diagonal is plotted to ease evaluation of differences between sessions, dots above this line indicate better performance in session 2.

Table 6. Significance testing for all assessments in Experiment 2 at two time points. * indicate significance at α = .05

Test	*Bias*	*Limits of Agreement*	Units	*r (p)*	*t (p)*	*df*	*Cohen's d*
Gap	-0.3	[-3.86, 3.24]	Log2 (ms)	.5 (<.01)*	0.85 (.39)	50	0.12
DichoticFM	-0.2	[-2.88, 2.47]	Log2 (hz)	.35 (.01)*	0.92 (.36)	50	0.12
DioticFM	-0.36	[-2.32, 1.58]	Log2 (hz)	.52 (<.01)*	2.13 (.03)*	50	0.29
TM	-0.19	[-1.77, 1.38]	M (dB)	.54(<.01)*	1.65 (.1)	50	0.23
SM	-0.28	[-2.03, 1.46]	M (dB)	.7 (<.01)*	1.18 (.24)	50	0.16
STM	0.12	[-1.73, 1.99]	M (dB)	.5 (<.01)*	-1.19 (.23)	50	-0.16
No-Notch	-1.06	[3.29, -5.41]	SMR (dB)	.49 (<.01)*	2.22 (.03)*	50	0.31
Notch	-0.45	[-8.84, 7.92]	SMR (dB)	.46 (<.01)*	-0.75 (.45)	49	-0.1
Co-located	-0.26	[-4.29, 3.77]	SMR (dB)	.38 (<.01)*	0.82 (.41)	50	0.11
Separated	-1.96	[-7.8, 3.88]	SMR (dB)	.44 (<.01)*	4.62 (<.01)*	50	0.64
SpatialR	1.7	[-6.59, 7.22]	SMR (dB)	.5 (.11)	-3.39 (<.01)*	50	-0.47

Headphone effects

Below we show the limits of agreement plots and the scatters for test performance divided by headphone used. In this case, the effects of session are collapsed since headphone choice was counterbalanced across participants and sessions.

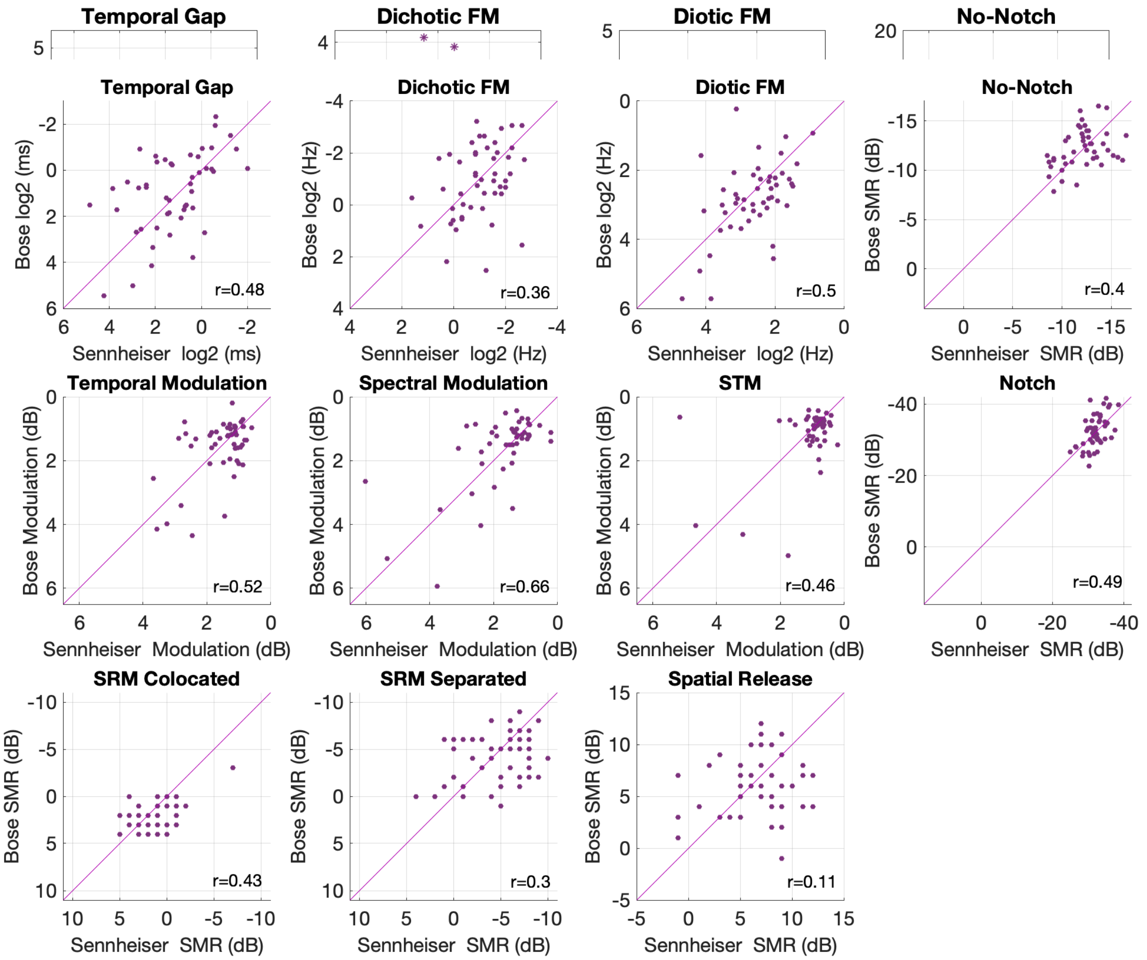


Figure 16. Scatter plots relating headphone types used for the 10 assessments in Experiment 2. All axes are oriented to show better performance values away from the origin. Correlations are indicated in the lower right of each panel. The diagonal is plotted to ease evaluation of differences between headphone types, dots above this line indicate better performance with active noise attenuation.

In agreement with our interpretation of a learning effect explaining the small but significant differences between sessions, there are no statistically significant differences between sessions performed with either headphone type. Once the chronological difference effect gets spread across the data for the headphone based split, the learning based bias disappears. Limits of agreement plots and scatters are very similar to both the session-based split of the data in Experiment 2 and the results of Experiment 1. This analysis further supports what is reported in the main manuscript regarding no effects of headphone type in silent conditions which was explored mainly through composite analysis.

Table 7. Significance testing for all assessments in Experiment 2 at two time points divided by headphone type used. * indicate significance at α = .05.

Test	*Bias*	*Limits of Agreement*	Units	*r (p)*	*t (p)*	*df*	*Cohen's d*
Gap	-0.26	[-3.83, 3.3]	Log2 (ms)	.48 (<.01)*	1.29 (.2)	50	0.18
DichoticFM	0.09	[-2.6, 2.79]	Log2 (hz)	.36 (.01)*	-0.37 (.7)	50	-0.05
DioticFM	0.24	[-1.78, 2.27]	Log2 (hz)	.5 (<.01)*	-1.27 (.21)	50	-0.17
TM	0.04	[-1.58, 1.66]	M (dB)	.52(<.01)*	-0.28 (.78)	50	-0.03
SM	-0.08	[-1.91, 1.74]	M (dB)	.66 (<.01)*	1.13 (.26)	50	0.15
STM	0.07	[-1.8, 1.95]	M (dB)	.46 (<.01)*	-0.17 (.86)	50	-0.02
No-Notch	-0.2	[-5.02, 4.6]	SMR (dB)	.4 (<.01)*	1.02 (.31)	50	0.14
Notch	-0.31	[-8.72, 8.09]	SMR (dB)	.49 (<.01)*	0.51 (.6)	49	0.07
Co-located	0.26	[-3.77, 4.29]	SMR (dB)	.43 (<.01)*	-0.82 (.41)	50	-0.11
Separated	0.56	[-6.36, 7.48]	SMR (dB)	.3 (.03)*	-1.12 (.26)	50	-0.15
SpatialR	-0.3	[-8.06, 7.46]	SMR (dB)	.11 (.46)	1.11 (.27)	50	0.15

Experiment 3 (noise; n=48)

This experiment is a replica of Experiment 2 except environmental noise recorded at one of UCR's local cafeterias was played through loudspeakers at about 70 dB SPL. Again a 1.5:1 step size ratio between the step-up and the step-down behaviors of the staircases of adaptive parameters was used, and half of our participants were tested using Sennheiser 285 Pro around the ear headphones on session 1 and the other half used Bose noise cancelling headphones. This was counterbalanced across sessions. Our aim with this manipulation was to see if our results on Experiments 1 & 2 generalized to conditions of environmental noise less ideal than a quiet sound booth or a testing room in a lab.

The main difference we find with these threshold estimates results and those reported in the main manuscript for the aggregated dataset come from the Notch Noise tasks. These tasks have wider limits of agreement than the previous experiments. An outlier analysis revealed these differences come mainly from outlying performance. A rejection criteria of ± 1.96 SD removes eight outliers from the Notch task which yields limits of agreement of [-5.36 to 3.7]. Six are rejected from the No-Notch task which yields limits of agreement of [-10.33 7.35]. Even when Experiment 3 produced more outlying performances in these tasks than Experiments 1 & 2, when they are removed, the limits of agreement become very similar to those reported in silent conditions.

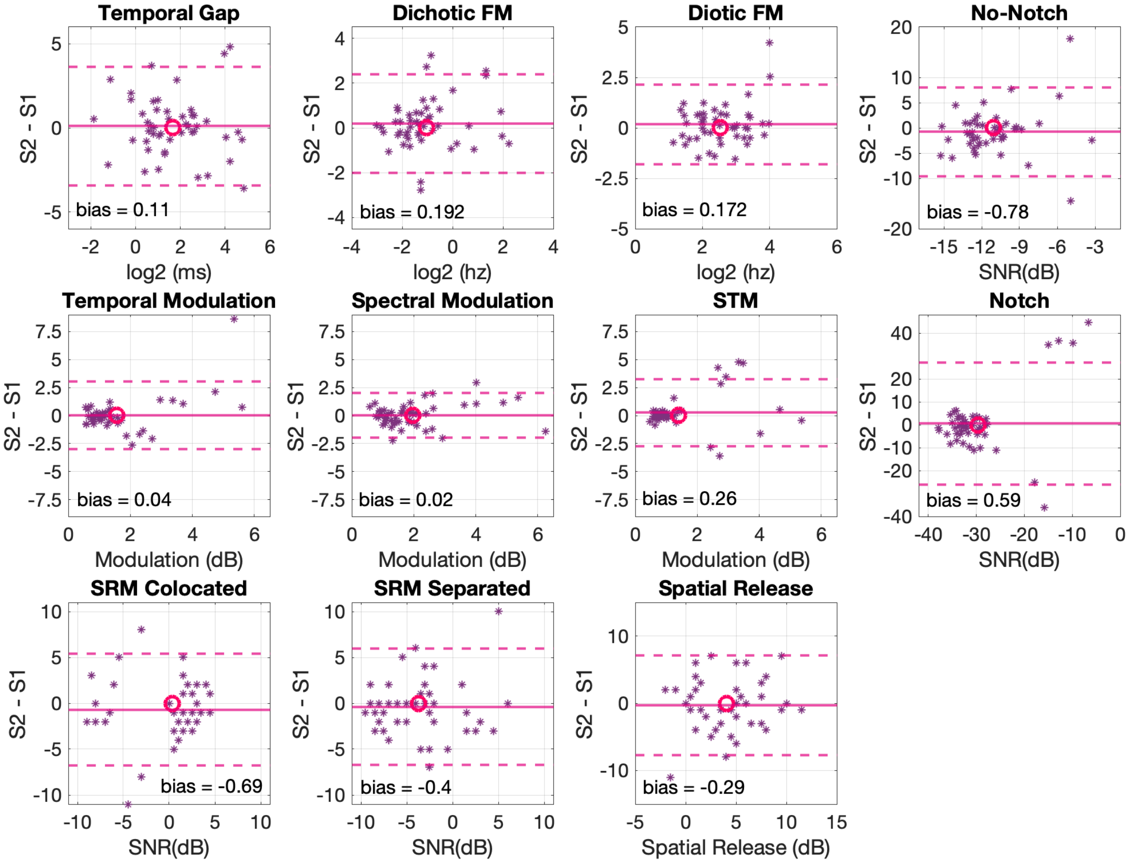


Figure 17. Limits of agreement of the estimated thresholds between sessions for all tests in Experiment 2. The solid lines indicate the mean difference between sessions. Dotted lines indicate the 95% limits of agreement. The red circle indicates the mean threshold for each test centered at cero difference between sessions. Solid lines below cero indicate better performance on session 2 (except the spatial release metric).

Another difference in this dataset was a subgroup of people that had very good performance on the SR Co-located condition. They provide an expanded region of performance that has a positive impact on the correlation. It is remarkable that these participants were able to perform so good even in the face of the extra masking provided by the external noise of our manipulation. The differences we observe in the co-located condition as well as the spatial release metric are largely explained by these outlying (good) performances.

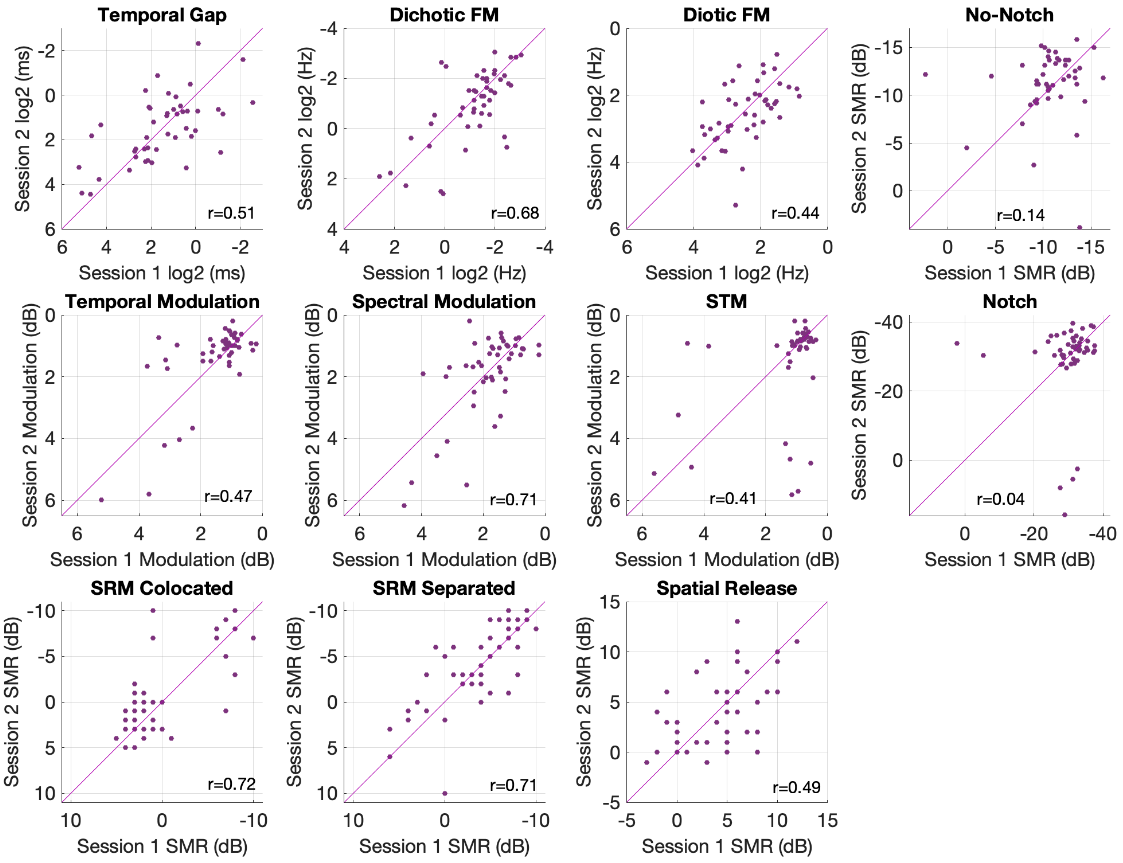


Figure 18. Scatter plots of Session 1 vs Session 2 for the 10 assessments in Experiment 3. All axes are oriented to show better performance values away from the origin. Correlations are indicated in the lower right of each panel. The diagonal is plotted to ease evaluation of differences between sessions, dots above this line indicate better performance in session 2.

Table 8. Significance testing for all assessments in Experiment 3 at two time points. * indicate significance at α = .05

Test	*Bias*	*Limits of Agreement*	Units	*r (p)*	*t (p)*	*df*	*Cohen’s d*
Gap	0.1	[-3.41, 3.63]	Log2 (ms)	.51 (<.01)*	-0.41 (.67)	47	-0.06
DichoticFM	-0.2	[-2.88, 2.47]	Log2 (hz)	.68 (.01)*	-1.18 (.24)	47	-0.17
DioticFM	-0.36	[-2.32, 1.58]	Log2 (hz)	.44 (<.01)*	-1.17 (.24)	47	-0.17
TM	0.03	[-2.99, 3.06]	M (dB)	.47(<.01)*	-0.16 (.87)	47	-0.02
SM	0.02	[-1.98, 2.03]	M (dB)	.71 (<.01)*	-0.16 (.87)	47	-0.02
STM	0.25	[-2.72, 3.24]	M (dB)	.41 (<.01)*	-1.17 (.24)	47	-0.17
No-Notch	-0.78	[8.05, -9.62]	SMR (dB)	.14 (.33)	1.2 (.23)	47	0.17
Notch	-0.45	[-26, 27.18]	SMR (dB)	.04 (.8)	-0.3 (.76)	47	-0.04
Co-located	-0.68	[-6.8, 5.43]	SMR (dB)	.72 (<.01)*	1.52 (.13)	47	0.22
Separated	-0.39	[-6.73, 5.94]	SMR (dB)	.71 (<.01)*	0.84 (.4)	47	0.12
SpatialR	-0.29	[-7.7, 7.11]	SMR (dB)	.49 (<.01)*	0.53 (.59)	47	0.07

Headphone effects

Below we show the limits of agreement plots and the scatters for test performance divided by headphone used. In this case, the effects of session are collapsed since headphone choice was counterbalanced across participants and sessions.

Table 9. Significance testing for all assessments in Experiment 2 at two time points divided by headphone type used. * indicate significance at α = .05

Test	*Bias*	*Limits of Agreement*	Units	*r (p)*	*t (p)*	*df*	*Cohen’s d*
Gap	-0.41	[-3.84, 3.01]	Log2 (ms)	.54 (<.01)*	1.65 (.1)	47	0.23
DichoticFM	0.02	[-2.21, 2.25]	Log2 (hz)	.67 (.01)*	-0.12 (.9)	47	-0.01
DioticFM	-0.02	[-2.03, 1.98]	Log2 (hz)	.42 (<.01)*	0.15 (.87)	47	-0.02
TM	-0.25	[-3.24, 2.72]	M (dB)	.46 (<.01)*	1.16 (.25)	47	0.16
SM	0.06	[-1.94, 2.07]	M (dB)	.71 (<.01)*	-0.44 (.65)	47	0.06
STM	-0.03	[-3.06, 2.99]	M (dB)	.39 (<.01)*	0.14 (.88)	47	0.02
No-Notch	0.75	[-8.1, 9.6]	SMR (dB)	.16 (.28)	-1.15 (.25)	47	-0.16
Notch	4.25	[-20.99, 29.5]	SMR (dB)	.23 (.12)	-2.28 (.02)*	47	-0.33
Co-located	-0.68	[-6.8, 5.43]	SMR (dB)	.73 (<.01)*	1.52 (.13)	47	0.22
Separated	-0.52	[-6.82, 5.78]	SMR (dB)	.72 (<.01)*	1.12 (.26)	47	0.16
SpatialR	-0.16	[-7.59, 7.25]	SMR (dB)	.49 (<.01)*	0.3 (.76)	47	0.04

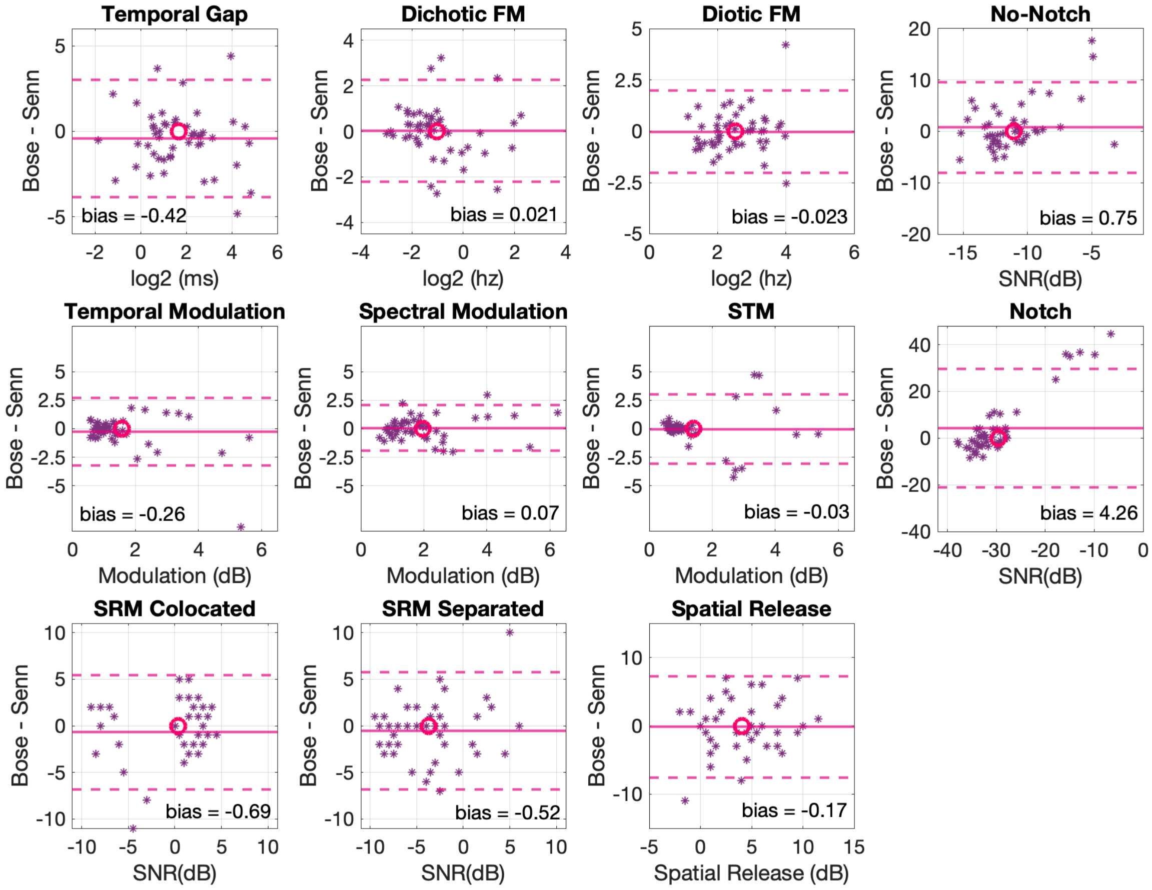


Figure 19. Limits of agreement of the estimated thresholds across headphone types in Experiment 3. The solid lines indicate the mean difference between headphone type. Dotted lines indicate the 95% limits of agreement. The red circle indicates the mean threshold for each test centered at cero difference between headphones. Solid lines below cero indicate better performance on the Bose headphones with active noise attenuation (except the spatial release metric).

This analysis revealed that the outlier performers flagged were using the Bose headphones. This is reflected more strongly in the Notch task where 6 participants are driving the statistically significant differences of about a third of a standard deviation in effect size that were observed in this task. After outlier rejection bias is reduced to -0.2 and the limits of agreement to [-9.52 to 9.11].

With the exceptions of the Notch tasks, the rest of the data shows more between-subject variability and thus the correlations improve slightly. Overall Experiment 3 replicates the findings of Experiments 1 & 2.

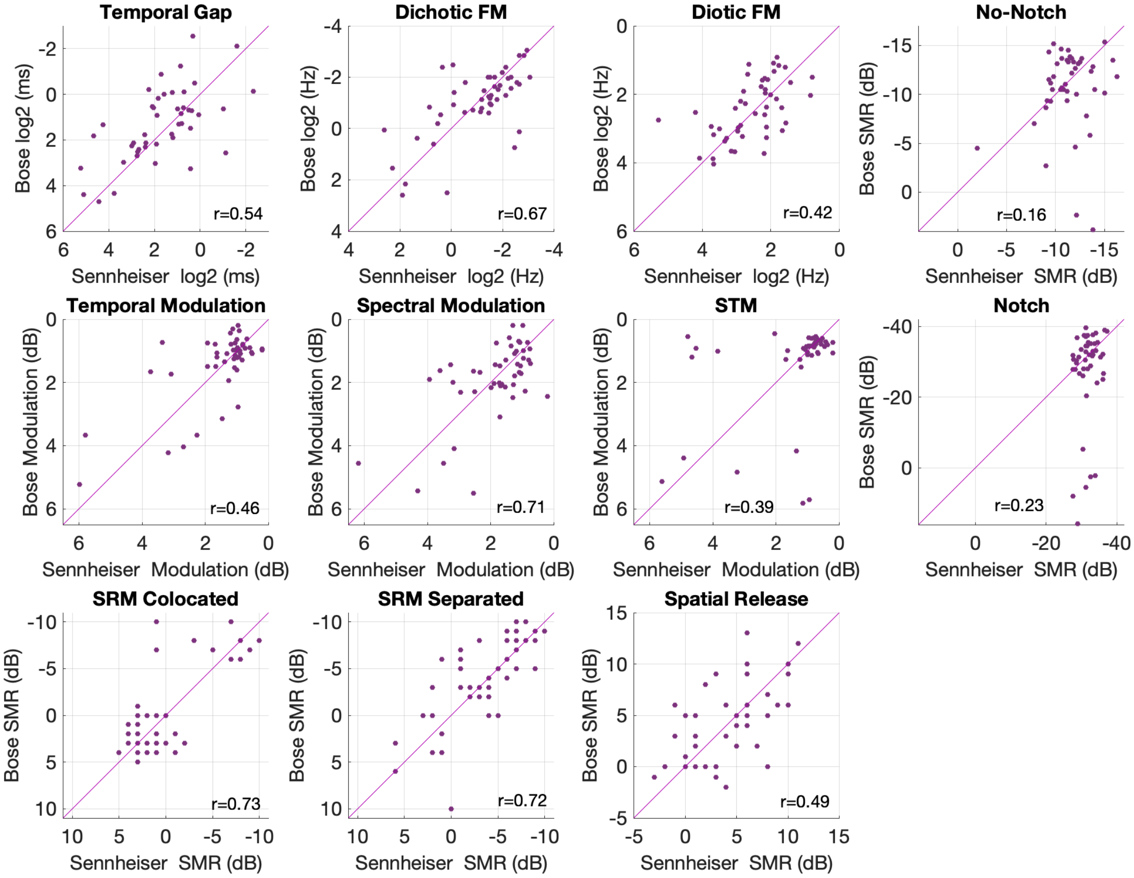


Figure 20. Scatter plots relating headphone types used for the 10 assessments in Experiment 2. All axes are oriented to show better performance values away from the origin. Correlations are indicated in the lower right of each panel. The diagonal is plotted to ease evaluation of differences between headphone types, dots above this line indicate better performance with active noise attenuation.

Adaptive Algorithm Effects

In this section, we provide details about individual progression trial by trial of the adaptive parameters selected for each task (except SR tasks which are on a fixed progressive track. Figures 21-28 show the individual and mean staircase progression for the adaptive PART assessments used. These figures can be used to further support the stability of our measures across Session and Experiment.

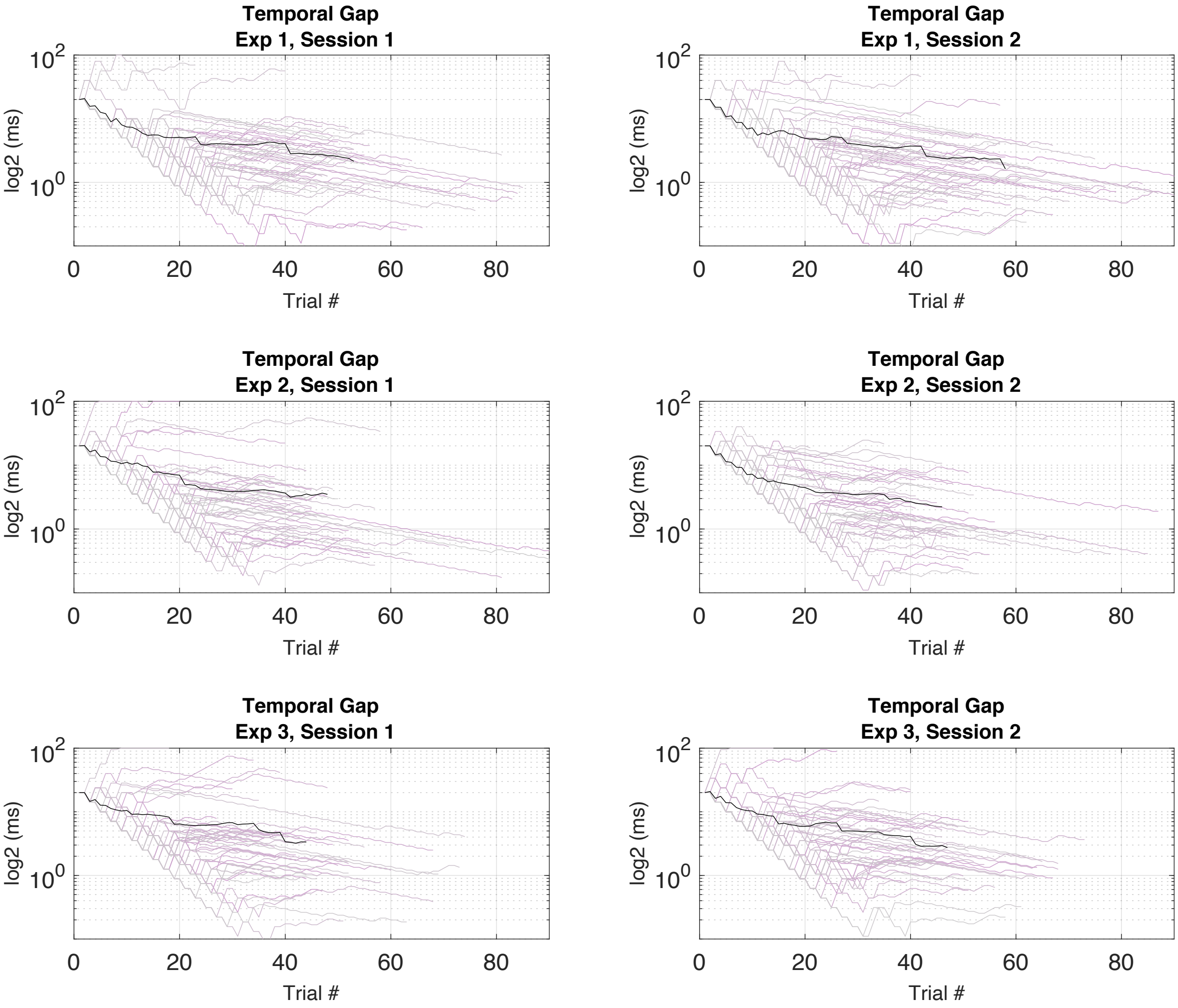
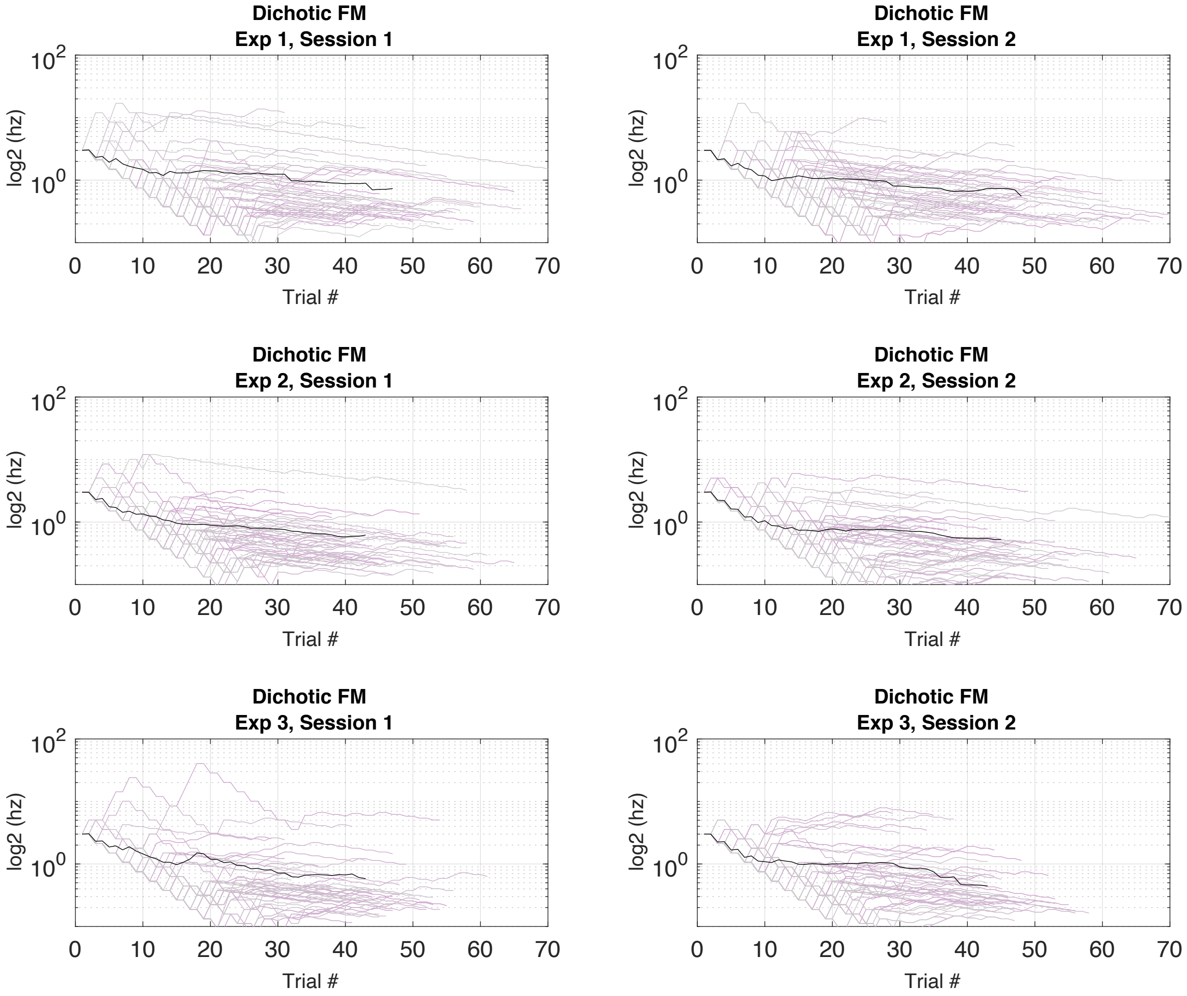


Figure 21. Adaptive staircase progression for the Temporal Gap detection task. Each row of sub-plots show a different Experiment. Panels on the left correspond to first sessions and panels on the right to second sessions. Each line represents a single participants going through the whole test. The black line represents mean performance.


Dichotic FM
Exp 1, Session 1

Dichotic FM
Exp 1, Session 2

Dichotic FM
Exp 2, Session 1

Dichotic FM
Exp 2, Session 2

Dichotic FM
Exp 3, Session 1

Dichotic FM
Exp 3, Session 2

Figure 22. Adaptive staircase progression for the Dichotic FM detection task. Each row of sub-plots show a different Experiment. Panels on the left correspond to first sessions and panels on the right to second sessions. Each line represents a single participants going through the whole test. The black line represents mean performance.

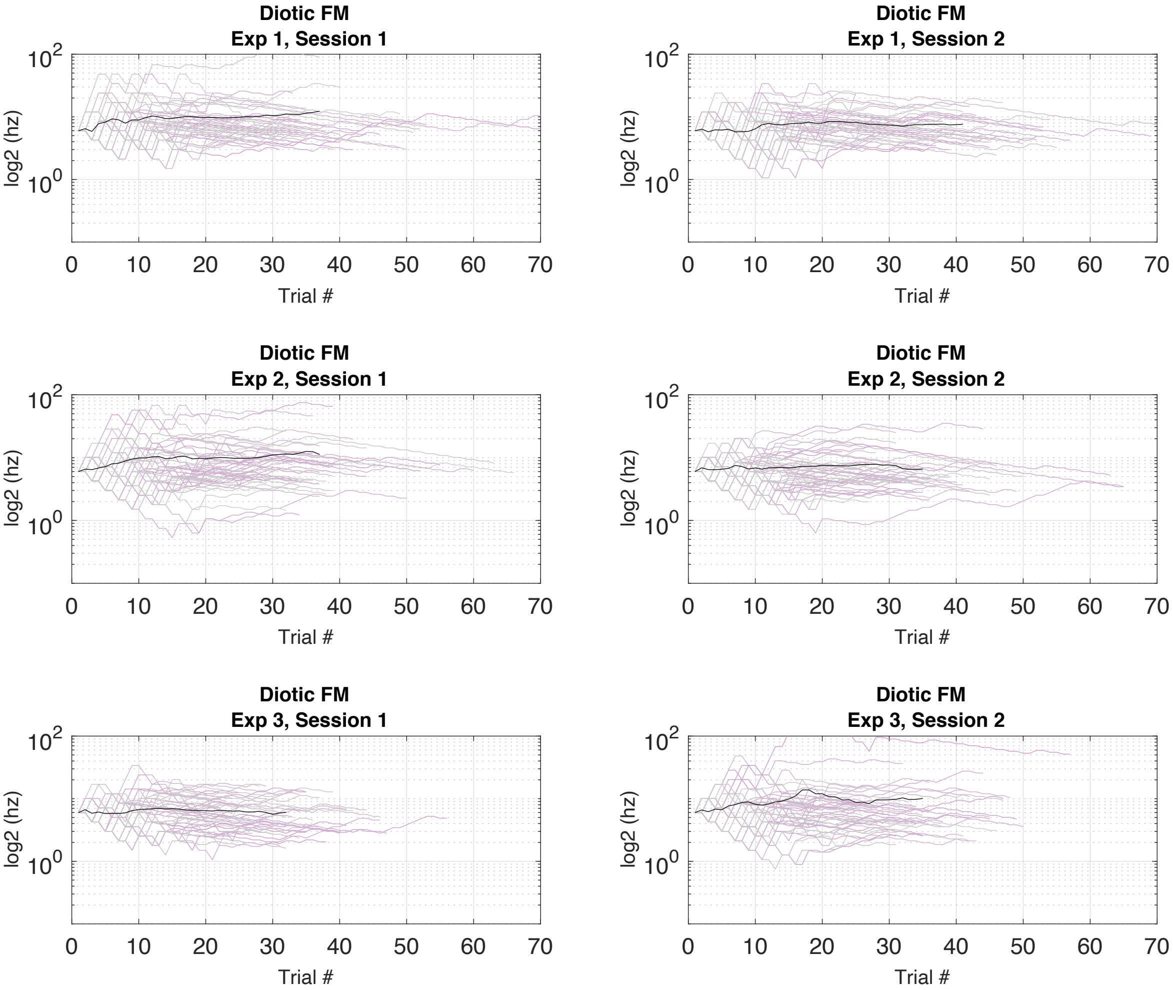


Figure 23. Adaptive staircase progression for the Diotic FM detection task. Each row of sub-plots show a different Experiment. Panels on the left correspond to first sessions and panels on the right to second sessions. Each line represents a single participants going through the whole test. The black line represents mean performance.

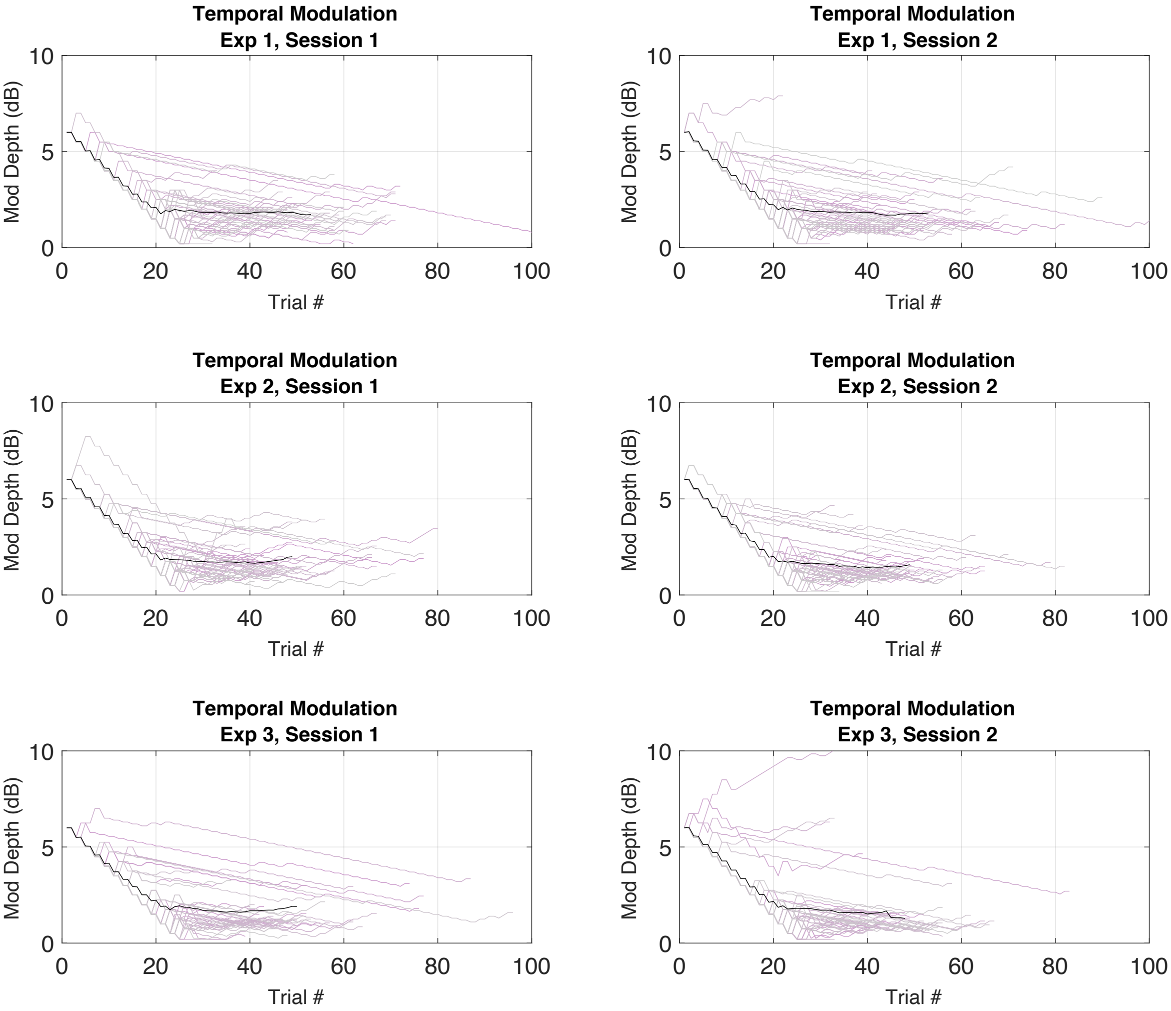


Figure 24. Adaptive staircase progression for the Temporal Modulation detection task. Each row of sub-plots show a different Experiment. Panels on the left correspond to first sessions and panels on the right to second sessions. Each line represents a single participants going through the whole test. The black line represents mean performance.

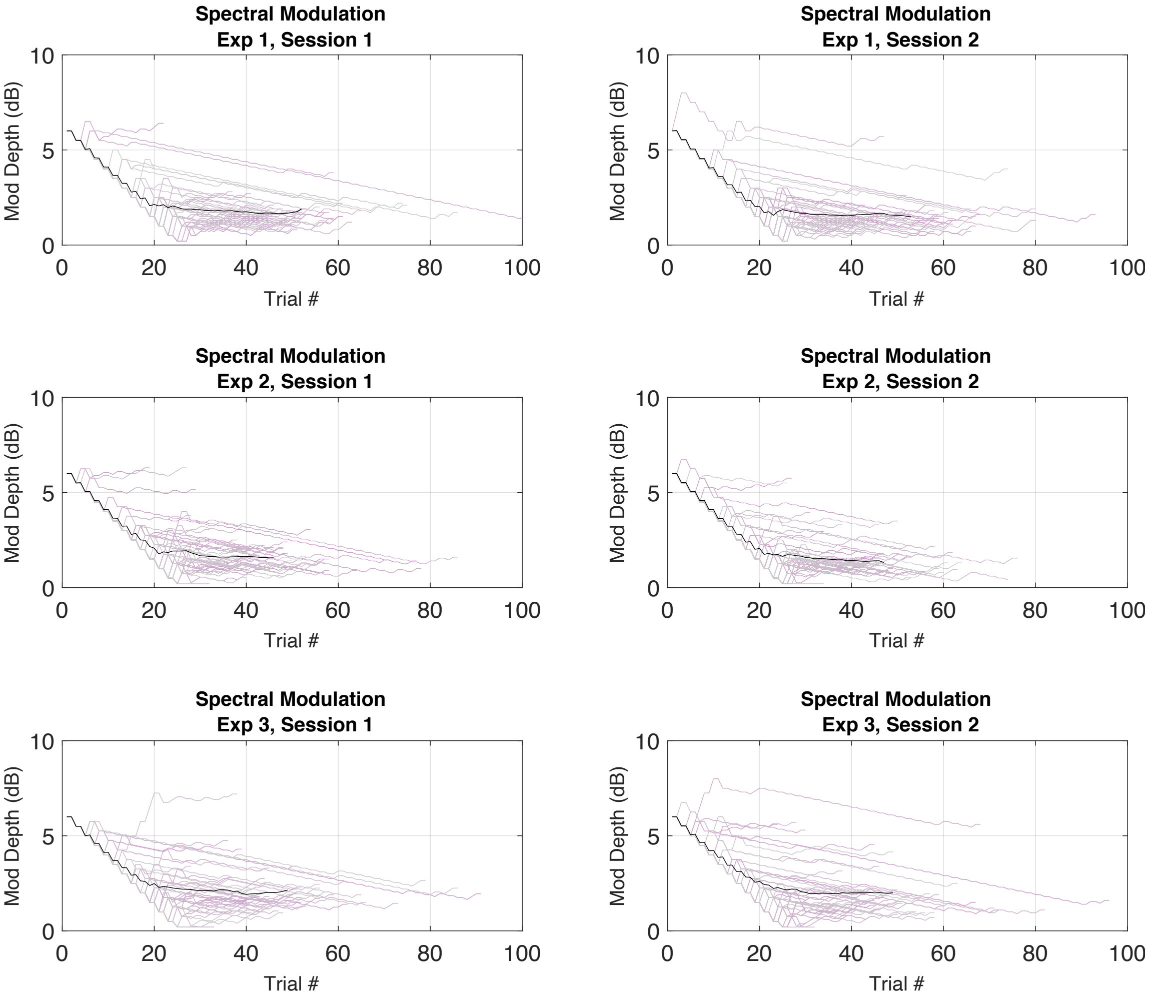


Figure 25. Adaptive staircase progression for the Spectral Modulation detection task. Each row of sub-plots show a different Experiment. Panels on the left correspond to first sessions and panels on the right to second sessions. Each line represents a single participants going through the whole test. The black line represents mean performance.

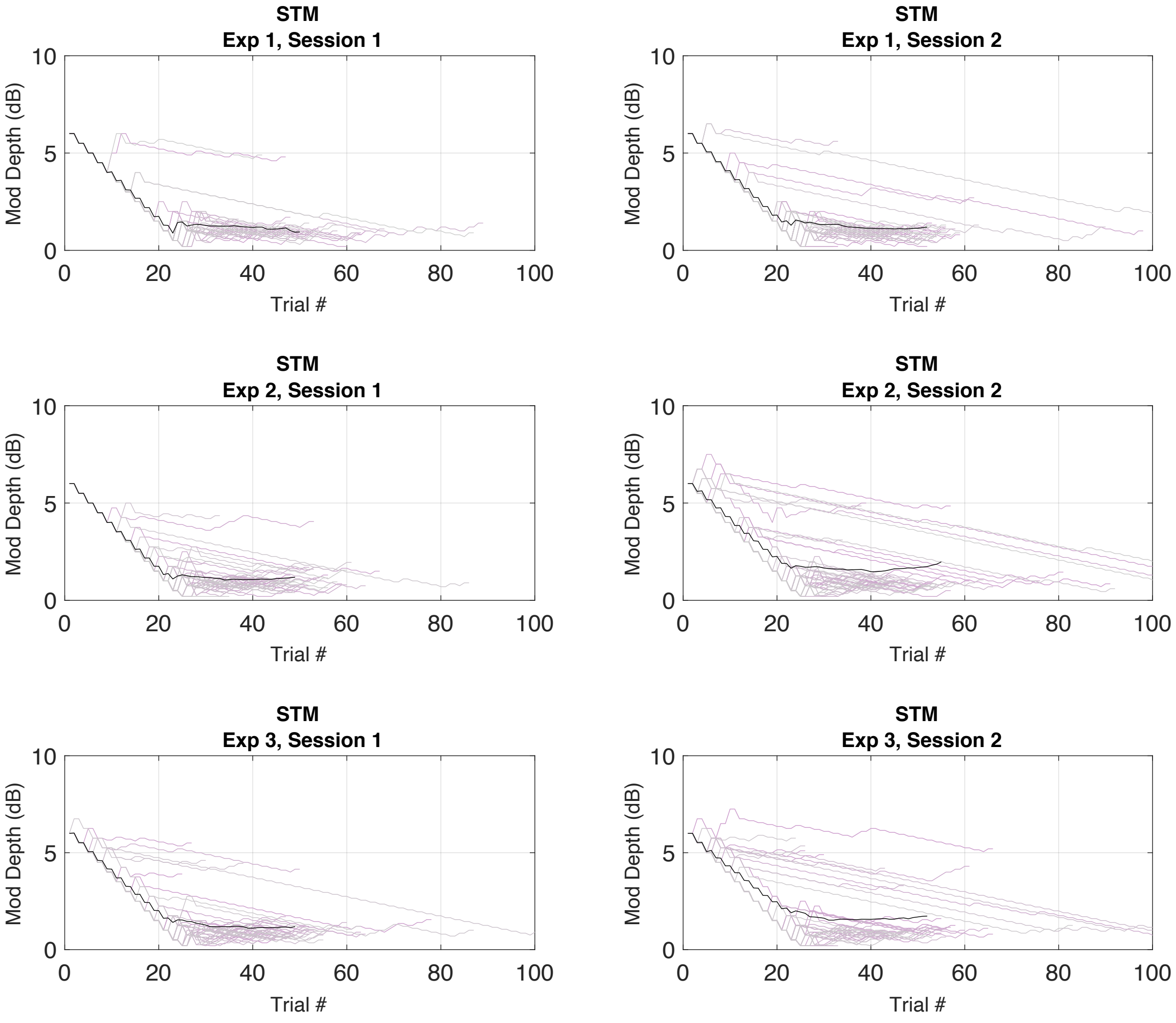


Figure 26. Adaptive staircase progression for the Spectro-Temporal Modulation detection task. Each row of sub-plots show a different Experiment. Panels on the left correspond to first sessions and panels on the right to second sessions. Each line represents a single participants going through the whole test. The black line represents mean performance.

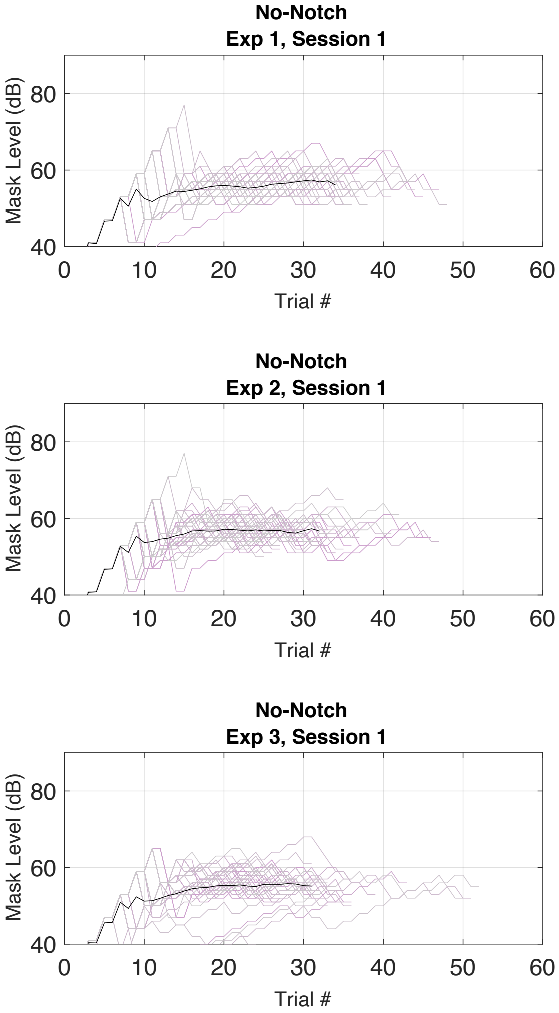


Figure 27. Adaptive staircase progression for the No-Notch 2 kHz tone detection task. Each row of sub-plots show a different Experiment. Panels on the left correspond to first sessions and panels on the right to second sessions. Each line represents a single participants going through the whole test. The black line represents mean performance.

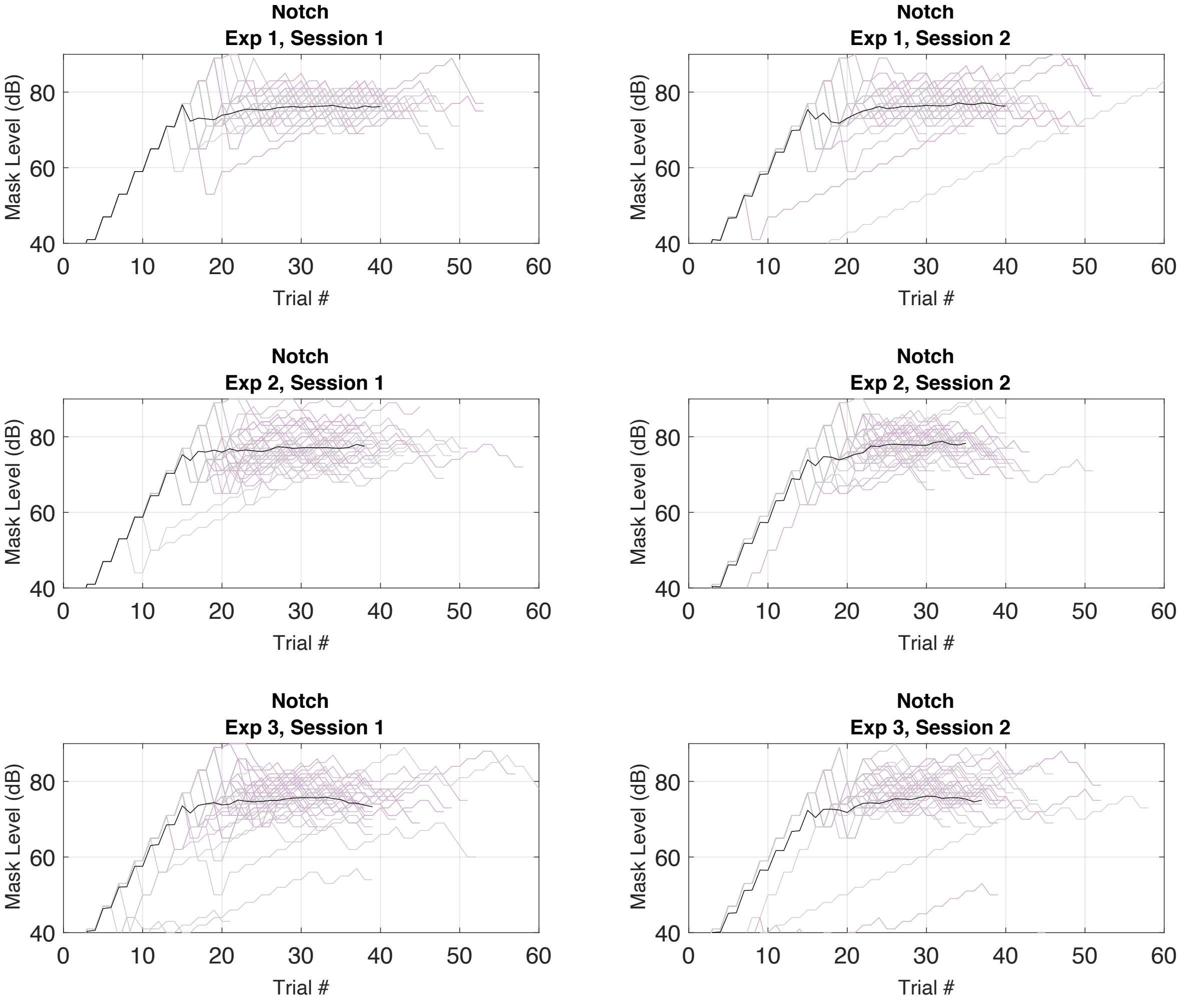


Figure 28. Adaptive staircase progression for the Notch 2 kHz tone detection task. Each row of sub-plots show a different Experiment. Panels on the left correspond to first sessions and panels on the right to second sessions. Each line represents a single participants going through the whole test. The black line represents mean performance.

As can be observed in the figures above, staircase data is consistent across both Session and Experiment. This analysis provides further support to the absence of experimental manipulation effects we reported in the main manuscript.

Trial Number Effects

To further test the consistency of our measures, analysis were conducted on the amount of trials needed to achieve the threshold estimate. Figure 29 shows the distribution of the number of trials done on each task for each Experiment. As can be observed here, the tasks behaved consistently across Experiment. Figure 30 shows the mean number of trials done per task per experiment, and the statistics for a one level ANOVA with the between subject factor Experiment (3 levels).

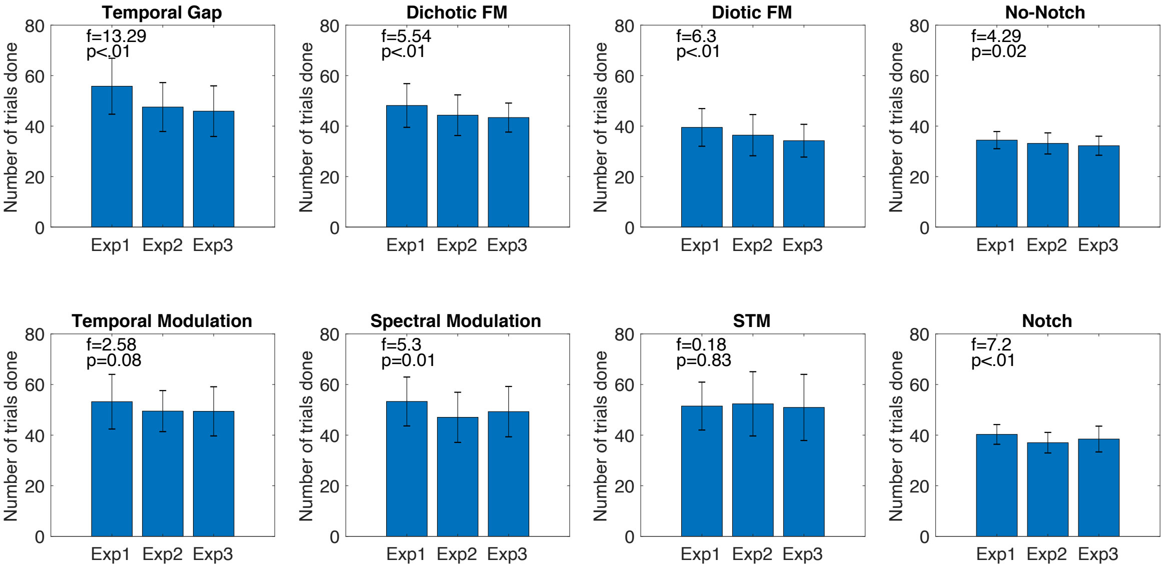


Figure 29. Shows mean and standard deviations of the number of trials presented per task for each Experiment. In addition, the statistics for a one-way ANOVA with the between-subject factor are displayed in the top of each graph.

To address the effects specific to the staircase a series of t-tests were conducted between the number of trials needed to achieve the threshold estimate in Experiment 1 and Experiment 2. These two experiments have the most similar conditions as well as the difference between step size ratios. Table 10 shows the results of the t-tests along with effect sizes. We found statistically significant differences or close to that in most of the tests with the 1.5:1 staircase needing less trials to finish. These differences are no more than 6 trials on average and sum up to 25.8 trials on average for the whole battery.

Table 10. Significance testing for the number of trials done for each assessment in Experiments 1 vs Experiment 2 * indicate significance at α = .05

Test	*Mean Difference (SD)*	*t (p)*	*df*	*Cohen's d*
Gap	6.2 trials (16)	4 (<.01)*	100	0.79
DichoticFM	3.6 trials (13.9)	2.32 (.02)*	100	0.46
DioticFM	3.6 trials (12.8)	1.98 (.05)	100	0.39
TM	3.3 trials (15)	1.96 (.052)	100	0.38
SM	3.4 trials (15.3)	3.23 (<.01)*	100	0.64
STM	1.7 trials (20.9)	-0.39 (.69)	100	-0.07
No-Notch	2 trials (7.8)	1.72 (.08)	100	0.34
Notch	2 trials (8.9)	4.15 (-.01)*	99	0.82

Overall, these results indicate the change in step-size from 2:1 to 1.5:1 resulted on staircases that were slightly more efficient to calculate a threshold estimate which itself did not differ across experiment in magnitude as shown by no significant independent samples t-tests between threshold estimates with a 3.96 SD filter. There were no significant differences between the number of trials done for any of the tests of Experiments 2 and 3 further supporting that the change in efficiency is relative to the change in step size ratio.

To conclude the supplementary presentation of the results, figure 30 shows the means and standard deviations obtained for each Experiment and Headphone type used with a ± 1.96 outlier rejection criteria. This figure is equivalent to figure 3 presented in the main manuscript except for the outlier rejection. It serves to confirm the stability of our measures without the noise of outlying performance. This are the values that were used for the second section of the results where our estimated thresholds were related to previous reports in the literature. In this figure it is easy to confirm the stability of our threshold estimates across experimental manipulations.

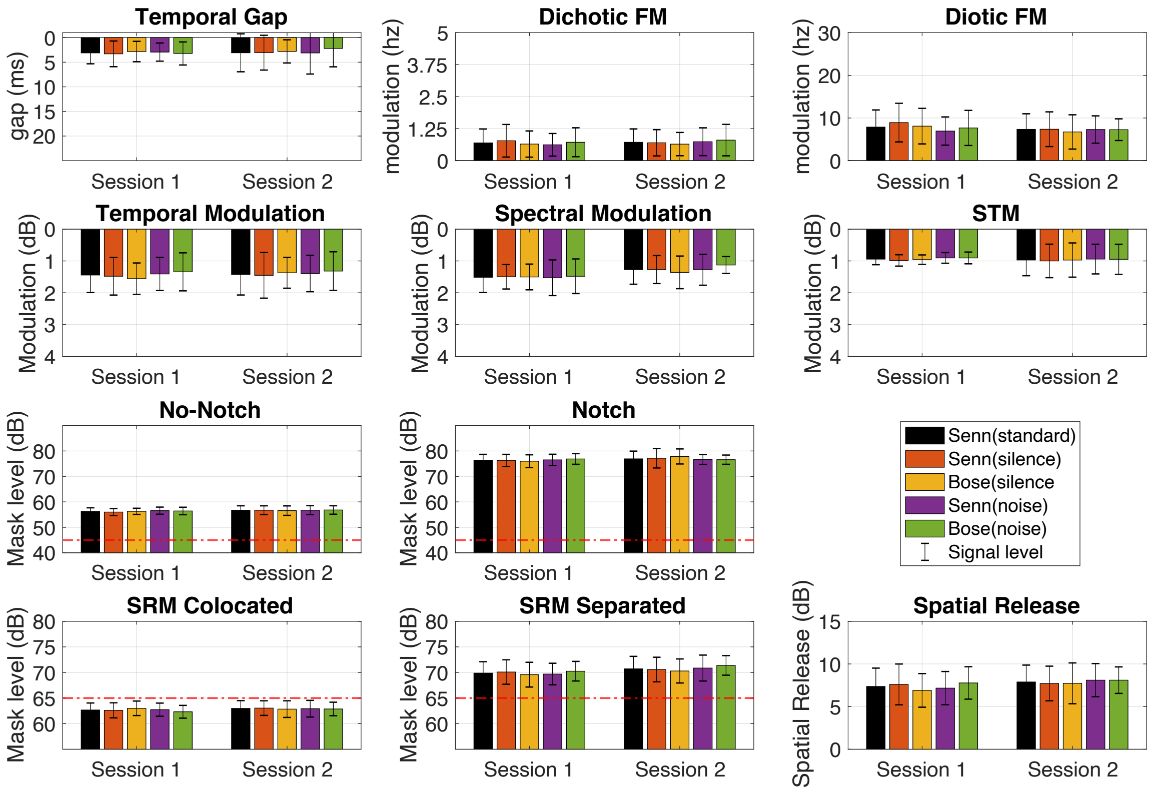


Figure 30. Mean thresholds and standard deviations obtained for each headphone used in each experiment in each test. The direction of the y-axes has been inverted when necessary so that better performance is always towards the top. A red dotted line indicates the level of the target in the target in competition tasks. Axis were kept the same as in figure 4 to facilitate comparison.