Comparison of loudness perception mismatch with visual presentation of two anechoic rooms

Manan Lamba[1], Simon Kersten[2], Stephan D. Ewert[1], Michael Vorländer[2], Birger Kollmeier[1]
[1] Medical Physics and Cluster of Excellence Hearing4all, CvO University Oldenburg, 26129 Oldenburg
[2] Institute for Hearing Technology and Acoustics, RWTH Aachen University, 52074 Aachen

Abstract
Many generations of researchers have tried to resolve the issue of loudness mismatch between the headphones and loudspeaker which occurs even when the same acoustical stimulus at the eardrum is presented. A preceding investigation on the loudness mismatch [5], reported a discrepancy in the mismatch across anechoic room conditions in Oldenburg and Aachen. In the present study, differences in room acoustics and in other non-acoustic influences as the potential causes for the observed discrepancy were tested. A headphone-only condition was employed where the visual presentation was varied using a head mounted display. To systematically separate potential acoustic difference between the rooms from the influence of the visually perceived room information, the headphone-equalized stimuli recorded with a head and torso simulator in the two anechoic rooms or diotic signals were presented acoustically while the visual presentation included both rooms or darkness. The visuals played negligible role in the perception of loudness of the recorded signals. But, the method of comparing the recorded signals was somewhat effective in generating the loudness mismatch, indicating towards the acoustical room effect in initiating the mentioned mismatch discrepancy. Hence, the findings provide encouragement for the future in exploring the mismatch with unreal sources in virtual auditory environments using headphones only.

Introduction
Ideally, one would assume that a sound signal at a certain loudness when presented through headphones and also through a loudspeaker would exert the same sound pressure level at the eardrum. This is also crucial for various everyday purposes including hearing aid fitting and for virtual reality applications, as well as for providing safe listening systems by avoiding abnormally high sound pressure levels. In reality, to produce the same loudness, a sound must be played louder by headphones than by loudspeaker. This phenomenon is known as the loudness mismatch. The loudness mismatch between the different acoustical presentation devices has been known since many decades and several partially successful attempts have been made to resolve it [1][2]. Though, the studies such as Völk and Fastl [3] and Brinkmann et al.[4], were able to eliminate the mismatch by the use of open headphones and the individual binaural synthesis.
Recently, Denk et al. [5] extensively examined the different binaural parameters and the effect of room conditions to reassess the loudness mismatch between headphones and loudspeakers. A comparative study across two sites [Aachen (AC) and Oldenburg (OL)] was performed which employed the respective anechoic room at each site and a group of subjects that performed the same experiments at both sites in addition to separate subjects at both sites. Additionally, one reasonably reverberant, echoic room was also included at each site. An unexplainable but statistically significant difference of approx. 3 dB between the loudness mismatch recorded at both the anechoic rooms, for all the narrow band stimuli and for diotic headphone playback was reported.

The anechoic rooms in consideration are quite different from each other. In the mentioned study, a probable reason for the difference in the loudness mismatch recorded among the anechoic rooms was thought to be these non-acoustical differences such as visual features in the room including the difference in number of loudspeakers in the corresponding rooms. A follow-up study was necessary to evaluate the influence of non-acoustical factors, in this case the visual cues presented to the subject. Meunier et al. [7] raised the issue of repeated repositioning of the headphones, during the transition from loudspeaker listening to headphone listening in direct comparison regarding the loudness mismatch. So, it was evident to undertake the loudness matching experiment where only headphones are used to play all the signals. Also, no investigation is known to the authors, in which the loudness mismatch experiment using the free field binaural stimuli, recorded at different locations and fed only through the headphones was performed. The exploration of the loudness mismatch using virtual reality is somewhat new and offers ample opportunity to learn unfamiliar aspects.

The different research questions addressed in this study are:

(a) Could the conventional loudness mismatch between headphones and loudspeaker also be reproduced by presenting the recorded free field signals through headphones?

(b) Does the discrepancy of the loudness mismatch observed in the two anechoic rooms in Denk et al. [5] also occurs with this method?

(c) What is the influence of the visual reinforcement i.e. the visual presentation of the respective rooms on the loudness mismatch?

Methods, procedure and set-up

Subjects performed a loudness matching experiment for the acoustical comparison for the two different sound signals. Both the signals were presented through the headphones only, one after the other. The subjects did not need to put the headphones on and off after each comparison, which was the case in the previous experiments. Computation of sound pressure levels at the eardrum was not necessary as the same headphones were used without any repositioning. As the transfer function from headphone to the eardrum did not change, the level differences were calculated based on the output signal without post-processing. Open coupling headphones (HD 650, Sennheiser, Wedemark, Germany) were used to present the stimuli, same as in the preceding study. An HpTF equalization was applied to the KEMAR recorded stimuli from respective rooms based on the mean of 8 repeated measurements. The phase of the mean was recreated assuming a minimum phase system [8].

A head mounted display (HTC Vive Pro Eye) was used to provide the visual cues i.e. 360^o photos of the anechoic rooms as well as a black screen. The subjects took part in only one experimental session. The 360^o photos of the specific anechoic rooms in Oldenburg and Aachen were taken with Xiaomi Mi Sphere Mijia 360 camera and Insta360 Pro 2- 360 VR Camera, respectively.

Nine different signals were used. Eight signals were one third octave band noises with centre frequencies at 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz, 12 kHz (referred to as tbn125, tbn250, tbn500 and so on). Ninth signal was a broadband noise with equal energy in each of 17 critical frequency bands as defined by Zwicker [9] in a frequency range between approx. 250 Hz and 4 kHz, i.e., the same lower and upper boundary frequency as the narrowband noises.The signals were 1 s in duration consisting of 20 ms long rise and fall ramps. These stimuli were binaurally recorded in free-field using a head and torso simulator mannequin, KEMAR 45BB-12 with anthropometric pinnae and low-noise ear simulators (G.R.A.S., Holte, Denmark) in the Oldenburg (LS_{OL}) and Aachen (LS_{AC}) anechoic rooms respectively. Also, the diotic stimuli i.e. the same signal presented to both the ears was used as an equivalent to the conventional headphone presentation for the loudness matching experiment. Taking note from the preceding studies which used the same resources, the focus was on the four signals as examined earlier [5][6].

1-up-1-down alternative forced choice paradigm [10] [11], implemented in the AFC toolbox [12], was employed in the loudness matching experiment. Verbal instructions about the experimental procedure were given to the subjects before the start of the experiment. During the whole procedure, the subjects were wearing the head mounted display. For each visual condition, the loudness matching experiment started randomly. A foot-switch with three buttons was used as the response device for louder stimuli selection. The left pedal of the foot-switch was assigned for the selection if the subjects perceived the first stimulus louder. Similarly, the right pedal was assigned the selection of the subsequently played signal if it was perceived louder. The pedal in the middle was used to proceed or continue the experiment after one stimuli combination with the acoustical comparison was done. During the course of the whole experiment, the subjects were not able to see the foot switch due to the use of virtual reality glasses. The step-size of the headphone playback level was lowered from the preliminary 10 to 5 dB and 1 dB following the first and second upper reversals respectively. The median value of the three upper and lower reversals of the headphone playback level throughout the measurement phase was gathered as the resultant equal-loudness level. A standard matching process needed between 12 and 20 comparisons until conjunction was attained, which took approx. 1-2 min for each acoustical condition and stimuli combination. Therefore, approximately 42 minutes on average was needed by the participants to complete each of the three visual conditions. Breaks were permitted after the completion of the experimental procedure for each visual condition. All possible conditions for each of the three visual conditions i.e. Stimulus x Acoustical comparison were performed in a random order. Table 1 shows the three different visual conditions as well as the acoustical comparisons along with nine different stimuli.

Table 1 – Different conditions examined in the experiment. Here, HP_{DI} denotes the diotic stimuli that is headphone equivalent of the conventional loudness matching experiment between headphone and loudspeaker. Also, LS_{OL} and LS_{AC} are the free field recorded signals from Oldenburg and Aachen anechoic rooms respectively, acting as loudspeaker equivalent of the matching experiment. In stimuli, tbn is terz band noise and uen stands for Unified excitation noise. The ones in bold letters have also been investigated previously by [5] & [6].

Visual Conditions	*Oldenburg anechoic, Aachen anechoic, Blind/Black screen*
Acoustical Comparisons	HP_{DI} *vs* LS_{OL}, HP_{DI} *vs* LS_{AC} , LS_{AC} *vs* LS_{OL}
Stimulus	*Third-octave-band noise: tbn125,* ***tbn250****, tbn500,* ***tbn1000****, tbn2000,* ***tbn4000****, tbn8000, tbn12000,* ***uen17***

Results

Thirteen self-reported normal hearing test subjects (10 male, 3 female) took part in the listening test. A three way analysis of variance (ANOVA) was performed, expecting the values of the level mismatch depending on the factors visual conditions, acoustical comparison and the stimuli. The visual conditions as in the 360^o pho-

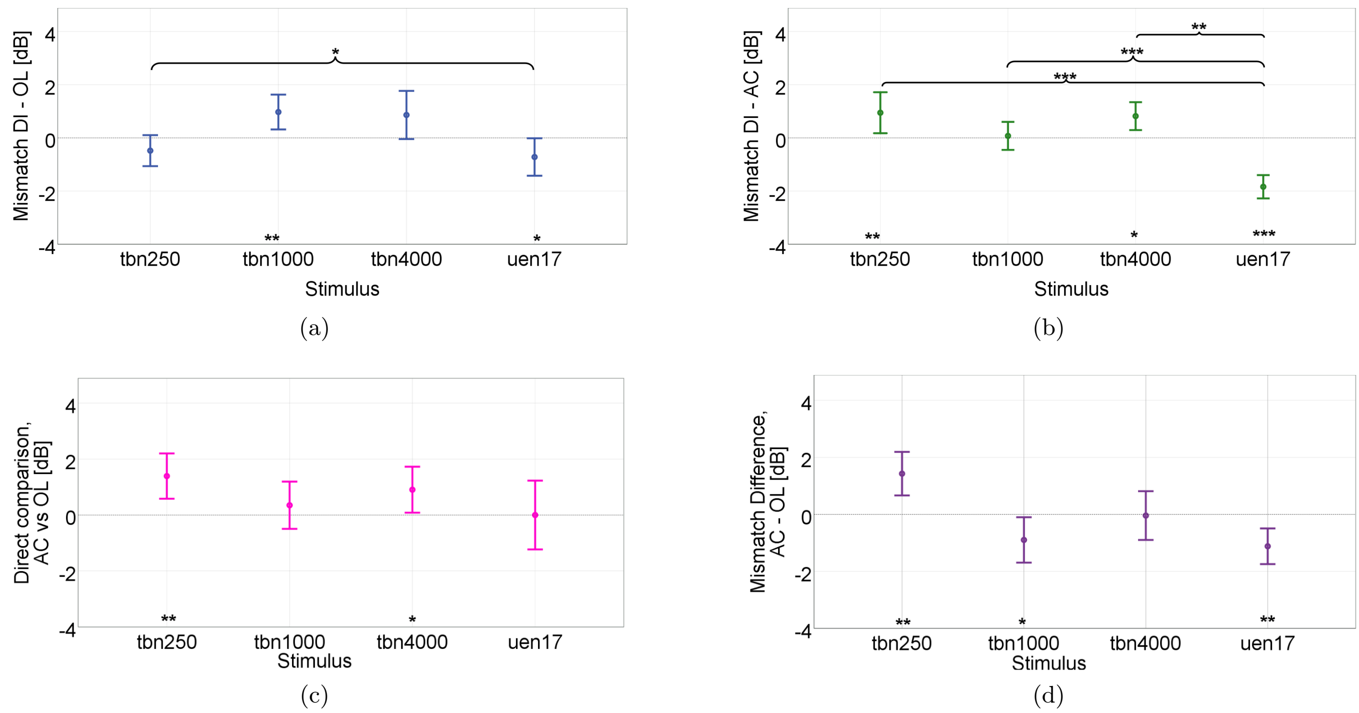

Figure 1 – The y- axis shows the mismatch i.e the level differences (in dB) for HP_{DI} vs LS_{OL} (fig. 1a), HP_{DI} vs LS_{AC} (fig. 1b), direct comparison: LS_{AC} vs LS_{OL} (fig. 1c) and mismatch difference (fig. 1d) respectively. The x-axis shows the different stimuli tested in each plot. The stars at the bottom indicate the significant level mismatch. Significant difference between different stimuli is indicated by perenthesis. The significance levels are shown by one, two or three asterisks for 0.05, 0.01 and 0.001 respectively.

tos of the two anechoic rooms or the black screen had no significant effect on the different levels at equal loudness. But, significant main effects of the acoustic comparison [$F(2,24) = 8.074$, $p < 0.01$] as well as the stimuli, $F[(8,96) = 7.852$, $p < 0.001]$ were observed. Also, significance was observed for an interaction between the acoustical comparison and stimuli, [$F(16,192) = 5.739$, $p < 0.001$]. Due to this reason, the data from all the subjects was averaged over the three visual conditions for a further analysis and T- tests with *alpha* equal to 0.05 were performed to determine the significant mismatch.
To prove association and relateability with the preceding investigation, the results have been shown for the *four* stimuli: tbn250, tbn1000, tbn4000 and uen17. Fig. 1 shows the mismatch during various comparisons as well as the difference. Plotted is the difference between Diotic stimuli and the loudspeaker recorded stimuli from two anechoic rooms, which is equivalent to the conventional comparison of loudness mismatch between headphones and loudspeaker has been shown in fig. 1a and 1b. Here, the asterisks at the bottom indicate the significant levels of the T-tests to determine whether there was a significant mismatch or not for a given stimulus. The brackets indicate significant differences between stimuli from the post hoc pair wise comparisons of the ANOVA with Bonferroni adjustments.
In the comparison of diotic versus Oldenburg anechoic room recordings (fig. 1a), no significant mismatch was observed for the low frequencies but the same was reproduceable for tbn1000. Also, an unexpected mismatch in the negative direction is observed for the broadband stimulus, uen17.

The diotic stimuli and the recorded stimuli in Aachen anechoic room comparison (fig. 1b) showed that a significant mismatch occurred for the lower frequency of tbn250 but was absent for tbn1000. For higher frequency of tbn4000 also, a mismatch was observed. Coincidently, in this case as well, a significant loudness mismatch was observed in the negative direction for the broadband stimulus, uen17.

The direct comparison (fig. 1c) of the binaurally recorded loudspeaker stimuli in both the locations yielded the significant mismatch for the lower as well as the higher frequencies of tbn250 and tbn4000 respectively. No mismatch was observed for the broadband stimulus which confirms the findings from the past studies [3][4].

The difference between the mismatch results for the different rooms (HP_{DI} - LS_{OL}) - (HP_{DI} - LS_{AC}) are shown in fig. 1d.

Summary and Conclusions

This investigation is an attempt to probe the loudness matching by employing the binaurally recorded free field stimuli and feeding it only through the headphones. The feedback indicated that some subjects perceived the sound source left or right direction instead from the exact front, for the broadband noise. This might be the cause due to the stimuli recording with a KEMAR and non-individualized headphone transfer functions (HpTFs). Additionally, the use of foot switch together with the head mounted display proved challenging for many participants, as there were reports of repeated unintentional and undesired response choices. Furthermore, different factors such as lack of training for the initial conditions as well as concentration loss might have lead to high variability in the listener responses. For the future studies, the above mentioned bottlenecks should be well managed. The use of a hand held controller as the response device for louder stimulus selection is also noteworthy. This would be advantageous in decreasing the uncertainty regarding the equipment and procedure among the subjects.

With respect to mentioned research questions in this study, the following conclusions could be drawn:

(a) The conventional loudness mismatch of around 6 dB between the headphone and loudspeaker presented signals could not be reproduced. Although, small but significant mismatch was observed for certain stimuli. The mismatch was observed mainly in the range of close to 1 dB. An unexpected mismatch in the negative direction i.e. -1 dB and approx. -2 dB was present for the comparison of diotic stimuli with loudspeaker recordings from Oldenburg and Aachen anechoic rooms respectively.

(b) The significant mismatch difference during the direct comparison of the binaural loudspeaker recordings as well as the mismatch the mismatch difference was between 2dB and -2dB. It could be noticed, that the discrepancy between the loudness mismatches at two different aneechoic rooms was also present by this method but not at the same level of approx. 3 dB as reported by Denk et al.[5].

(c) Lastly, the visual conditions had no effect on the recorded levels at equal loudness.

This method needs to be examined further. It opens a new dimension towards the virtual auditory scenes and also a step further in understanding the origin of the loudness mismatch. However, more data in combination with utilization of individual transfer functions is necessary to decide whether this method is beneficial or not.

Acknowledgements

We would like to thank all the subjects who took part voluntarily. Kristin Ohlmann is acknowledged for her suggestions and assistance during the implementation. This study was funded by the Deutsche Forschungsgemeinschaft (DFG) - Projektnummer 352015383 - SFB 1330 HAPPAA A4.

References

[1] Beranek, L. L. (1949). Acoustic measurements. New York: John Wiley & Sons.

[2] Killion, M. C. (1978). Revised estimate of minimum audible pressure: where is the "missing 6 dB"? J. Acoust. Soc. Am. 63, 1501-1508.

[3] Völk, F., and Fastl, H. (2011). "Locating the missing 6 dB by loudness calibration of binaural synthesis" in Audio engineering society convention. Vol. 131; October 2011. New York, NY, USA: Audio Engineering Society, 1-12.

[4] Brinkmann, F., Lindau, A., and Weinzierl, S. (2017). On the authenticity of individual dynamic binaural synthesis. J. Acoust. Soc. Am. 142, 1784-1795. doi: 10.1121/1.5005606

[5] Denk F, Kohnen M, Llorca-Bofi J, Vorländer M and Kollmeier B (2021) The "Missing 6 dB" Revisited: Influence of Room Acoustics and Binaural Parameters on the Loudness Mismatch Between Headphones and Loudspeakers. Front. Psychol. 12:623670. doi: 10.3389/fpsyg.2021.623670

[6] Kohnen, M., Llorca-BofÃŋ, J., Denk, F., Vorländer, M., and Kollmeier, B. (2019). Loudness in differeeadphone reproduction: Is there a mismatch even after careful equalization? Proceedings of the 23rd International Congress on Acoustics, 9-13th September, Aachen, Germany.

[7] Meunier, S., Chatron, J., and Bonetto-Lemaire, T. (2020). "The Missing 6 dB: Perceptual or Procedural Effect?" in Proceedings of the Forum Acusticum; December 2020; Lyon, France, 3039-3043.

[8] Norcross et al., "Inverse filtering design using a minimal-phase target function from regularization," AES Convention 121, 2006.

[9] Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J. Acoust. Soc. Am. 33:248. doi: 10.1121/1.1908630

[10] Kollmeier, B., Gilkey, R. H., and Sieben, U. K. (1988). Adaptive staircase techniques in psychoacoustics: a comparison of human data and a mathematical model. J. Acoust. Soc. Am. 83, 1852-1862. doi: 10.1121/1.396521

[11] Levitt, H. (1971). Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467-477. doi: 10.1121/1.1912375

[12] Ewert, S. D. (2013). "AFC - A modular framework for running psychoacoustic experiments and computational perception models." in Fortschritte der Akustik - DAGA; March 2013; Meran