Audio Engineering Society

Convention Paper 9855

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

A Study of Listener Bass and Loudness Preferences over Loudspeakers and Headphones

Elisabeth McMullin[1]

[1]*Samsung Research America, Audio Lab, Valencia, CA 91355, USA*

Correspondence should be addressed to Elisabeth McMullin (e.mcmullin@samsung.com)

ABSTRACT

In order to study listener bass and loudness preferences over loudspeakers and headphones, a series of experiments using a method of adjustment were run. Twenty listeners rated their opinion of the songs used in testing and then completed two training tasks to assess their ability to adjust bass and loudness levels consistently. Listeners then completed separate experiments in which they adjusted bass and loudness levels to their preference over loudspeakers and over headphones. The results indicate that listeners had greater difficulty adjusting bass and loudness levels consistently over headphones than over loudspeakers. On average listeners adjusted the bass 1 dB higher and the loudness 2 dB higher over loudspeakers than over headphones. Other interactions explored include song, listener training, and hearing ability.

1 Introduction

Bass and loudness are undoubtedly the two simplest and most vital aspects of a consumer's conception of audio quality. With these two aspects come a series of stereotypes about listening behaviors, musical preferences, and demographic attributes. The goal of this research is to further understand listener perception of the relationship of these fundamental audio aspects across playback methods.

Previous research [1][2], explored similar questions of listeners' preferred bass levels. A shortcoming in these studies, as noted by the authors, was that the loudness was not compensated for when the listener turned up the bass. Because of this, it was impossible to isolate the variables of bass level and volume, and some less trained listeners may have turned up the bass when they actually wanted more level. In a later paper [3], the authors ran an in-ear headphone bass preference test using loudness normalization and found that without loudness normalization, listeners turned up the levels of a bass shelf an average of 2 dB more than with normalization engaged.

Several studies have been published about preferred listening level, but few have focused on consumer stereo listening environments. Dash et al. [4] found the mean preferred listening level for televisions playing music to be around 62.5 dBA. In contrast, Benjamin and Crockett [5] found that in an automotive environment, the mean music listening level in a stationary car without the engine running was around 73.7 dBA.

In regards to headphones, there have been many papers focused on the playback levels of portable players through insert and intra-concha earphones with and without background noise. In Worthington et al. [6], researchers found that in a quiet environment, the mean listener playback level measured at the ear drum was 71.9 dBA using a listener's own music and playback device. There was a large amount of variance across listeners and the listener preferred levels ranged dramatically from 53.4 dBA to 89.1 dBA.

King et al. [7] explored the differences between the results of sound engineers mixing different genres of songs over headphones and loudspeakers. His team found that when mixing classical and rock music over loudspeakers, there was much less variance between mixers than when they used headphones. For different styles of music, mixers tended to mix louder over different playback methods, but there was no consistent trend across genres.

Notably, there are dramatic differences between the nature of playback of stereo recordings over headphones and stereo loudspeakers. When listening to a stereo recording over loudspeakers, there is always crosstalk in which both the contralateral and ipsilateral ear receive the audio signal from each of the two speakers with a slight time delay. Additionally, the sensations of bass between the two playback methods are inherently different; over headphones, bass is largely only heard, while over loudspeakers, bass can be felt in the body. This study does not try to account for these major differences by using binaural recordings of loudspeakers, introducing artificial crosstalk, or using haptic feedback devices in the headphone tests. Measures were taken, however, to ensure that the headphones and loudspeakers start from a similar spectral and loudness position.

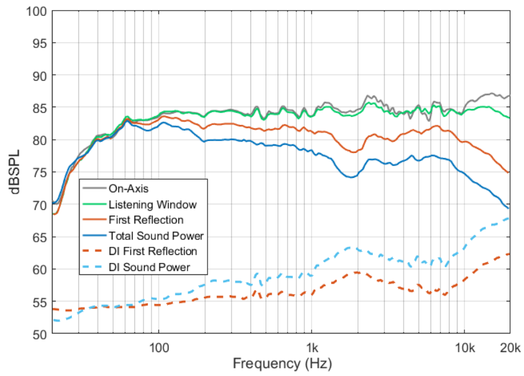
In this paper, bass and loudness levels are evaluated separately, without any level normalization in the bass adjustment session or bass compensation in the loudness tests. While there is likely a recursive interaction between these variables, the author feels it is important to evaluate their effects separately before running tests that attempt to compensate for their interactions. Hopefully these questions can be addressed in future research.

2 Experiment Setup

2.1 Loudspeakers, Headphones and Calibration

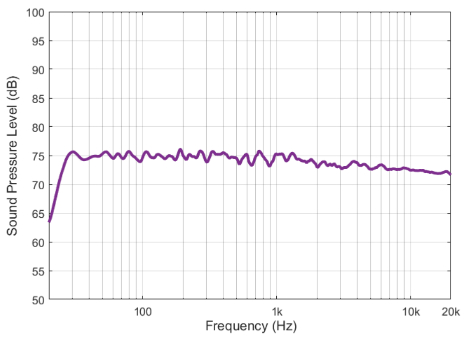
For the loudspeaker portion of the experiments, a pair of high-quality 3-way Revel Ultima Studio2 loudspeakers were selected. The loudspeakers were configured in a $\pm 30^{\circ}$ stereo configuration in the Samsung Audio Lab Small Listening Room [8] which was configured as an ITU-R BS.1116-1 listening room [9]. The listener was positioned 2.45 m back from each loudspeaker.

Fig. 1: Revel Ultima Studio2 "Spin-o-rama" measurement made anechoically in a 4-pi chamber per [10]



Eight GRAS 40BD microphones were placed around the seating position with 0.4 m between each mic. A spatial average was then made of the in-room response of each loudspeaker. Using the Samsung Audio Measurement System (SAMSLab), an AutoEQ algorithm was run to match the spatial average curve to a flat target curve with a high-frequency tilt above 3kHz (first-order high shelf with -3 dB gain). This algorithm calculates IIR biquad filters that will modify a measured curve to best meet a target curve. Care was taken when using this method to avoid adding excessively large peaks in the bass and to favor dips to avoid adding audible distortion. Excluding the bass, the target curve closely resembles the preferred in-room curve above mentioned in [11]. The filters calculated by the SAMSLab AutoEQ were applied to the audio signal using IIR filters loaded into a BSS BLU-160 signal processor. After calibration, a GRAS KEMAR manikin fitted with RA0045-S1 ear simulators and large anthropometric

Fig. 2: Eight mic spatial average with 1/6-oct smoothing of the left and right loudspeakers at the listening position after applying AutoEQ filters



pinnae was placed in the listening position to document the DRP (drum reference point) of each loudspeaker.

For the headphones tests, open-backed circumaural Beyerdynamic DT-990 Pro headphones were utilized. Measurements of the headphones were made using the same KEMAR manikin used to document the room measurement. The headphones were reseated five times on the manikin and frequency response measurements were made. The measurements were highly repeatable, but to account for the minor variations, an energy average was made of measurements. Using the SAMSLab AutoEQ, a set of IIR filters was calculated which could transform the current headphone curve to meet a version of the “2013 Harman Target Curve” [1] with flattened bass. These filters were imported into the custom software used in the experiments.

Fig. 3: Average of five reseats of DT-990 Pro headphones measured on a KEMAR head with ear simulators. Left Ear (black) Right Ear (red)

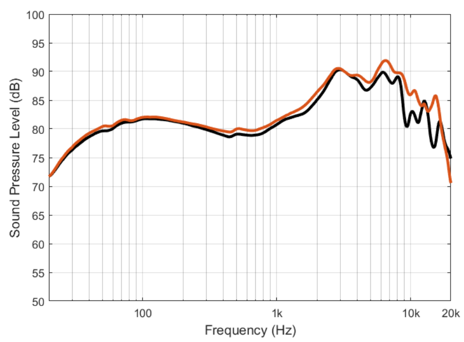
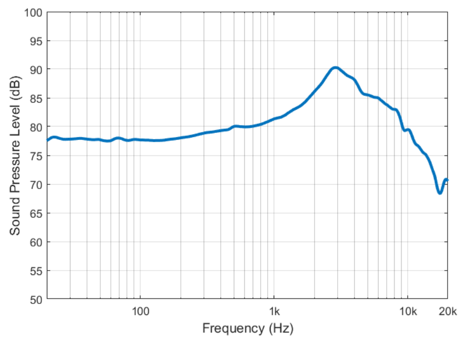


Fig. 4: Average of the left and right channels of a Beyerdynamic DT-990 Pro headphone after eq was applied to meet the "2013 Harman Target"

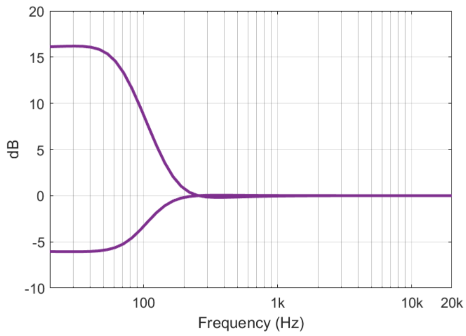


2.2 Level-Matching and Adjustment Ranges

Using a KEMAR manikin equipped with ear simulators, both the headphones and loudspeakers were level-matched to 70 dBA using -18 dBFS (ITU-1770-4) uncorrelated pink noise. This level was used as the default playback level for the bass and song rating portions of the experiment. Upon initial observations, the headphones sounded quieter than the loudspeakers at the same level. Many authors have found a similar effect, as Bech and Zacharov note [12] "In practice, this means that if headphones are calibrated to the same physical level as a free-field source, the headphone will be perceived to sound quieter than the free-field source." They go on to note that there is currently no concrete guidance on exact calibration levels available. In an attempt to compensate for this issue, KEMAR measurements of both the headphones and loudspeakers were compared and a 2.5 dB compensation gain was added to the headphones in order to compensate for the crosstalk experienced in stereo loudspeaker listening. Note that in the results, the headphone levels refer to the level without this added compensation gain.

For loudness adjustment tasks, the in-room level ranged from 57.6 dBA to 81.5 dBA using -18 dBFS uncorrelated pink noise. For bass adjustment tasks, listeners adjusted a second-order bass shelving filter set with a corner frequency set at 105 Hz with a range of adjustment from -6 dB up to +16 dB as shown in figure 5. This corner frequency was chosen because it is near the crossover frequency for many consumer audio products (particularly soundbars), it allows for comparison to

Fig. 5: The range of adjustment of the bass shelving filter for the bass adjustment tasks



prior experiments [1][2], and it allows for clearly audible changes over both headphones and loudspeakers. At lower crossover frequencies, the gain changes were more difficult to distinguish over headphones than over loudspeakers, which is likely due to the loss of full body bass sensations.

2.3 Program Material

Six programs were selected for their bass extension, genre representation, spectral balance, and relevance to contemporary music. All songs chosen had been produced within the past ten years, excluding the Steely Dan track, which was included to allow for comparison to previous studies. All songs were level matched to -18 dBFS (ITU-R BS 1770-4). Before each headphone or loudspeaker testing session, listeners rated the song clips using a 5-point Likert scale (1-Strongly Dislike, 2-Dislike, 3-Neutral, 4-Like, 5-Strongly Like) to rate their opinion of the musical content without regard to audio quality.

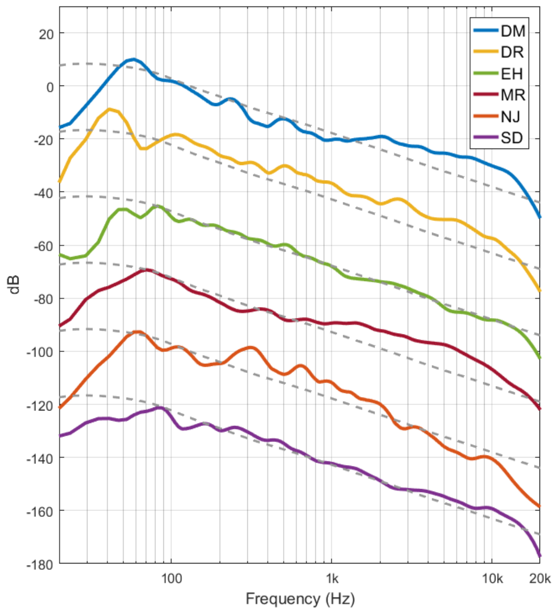
2.4 Listeners

Twenty subjects were included in the study, all of whom were employees or contractors of Samsung Research America. They ranged in age from 23 to 59 with a median age of 41.65 (SD = 11.26). Three of the listeners were female and 17 were male. All were tested for normal audiometric hearing and their audiograms were documented for use in the analysis. Listeners were categorized by their level of high-frequency loss on three levels and were placed in a category based on loss in one or both ears: none (less than 25 dBHL loss), mild loss (25 dB-40 dBHL), moderate loss (greater than 40

Table 1: Audio Programs used in testing

Artist / Genre	Song/Album/Label
DeadMau5 (DM) / Dance Electronic with Male Vocal	Ghosts n Stuff / For Lack of a Better Name / mau5trap 2009 CD
Drake (DR) / Hip Hop with Male Vocal	One Dance / Views / Cash Money 2016 CD
Emmylou Harris and Rodney Crowell (EH) / Country with Mixed Vocals	Black Caffeine / Old Yellow Road / Nonesuch 2013 CD
Mark Ronson and Bruno Mars (MR) / Pop Funk with Male Vocal	Uptown Funk / Uptown Special/ RCA 2014 CD
Norah Jones (NJ) / Jazz with Female Vocal	It's a Wonderful Time for Love/ Day Breaks / Blue Note 2016 CD
Steely Dan (SD) / Jazz Rock with Male Vocal	Cousin Dupree / Two Against Nature / Giant 2000 CD

Fig. 6: Average left and right channel power spectral density of six songs used with 1/3-octave smoothing plotted over 1.5-octave smoothed pink noise. Each curve is offset by 25 dB from the curve above it.



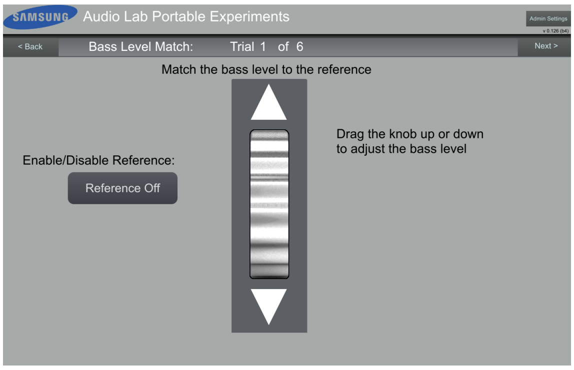
dBHL loss). One listener had moderate loss bilaterally and five listeners showed signs of mild loss unilaterally. Considering the percentage of listeners with hearing loss 4kHz and below in the US population is around 20%[13] and that our audiometry tests include up to 8 kHz, this percentage of listeners with mild loss is similar to the larger population. Twelve of the listeners were considered trained listeners based on their consistency in past listening tests based on their F_l [12] scores from previous experiments. Half of the listeners participating in the study began the test sessions over loudspeakers while the other half listened over headphones to account for potential training effects.

2.5 Test Design

All tests were administered on a Samsung Galaxy TabPro S tablet running custom software built using Max 7. The tablet was connected to an external audio interface for both the headphone and loudspeaker tests. The custom software recorded basic demographic information about each listener, randomized the playback order of songs and starting levels for the bass and loudness tasks, and recorded results into a SQLite database.

Listeners completed five tasks in each testing session. The first task was a song rating task. Listeners heard each of six test songs in a randomized order and rated them on a five-point Likert scale based on their opinion of the musical content. Following the rating task, listeners completed two six-trial training tasks: one for loudness and one for bass. In these tasks, listeners used a method of adjustment procedure to increment the bass or loudness level of the audio using a virtual "infinity thumbwheel," which had visual detents but no indicators of position or level. The listener swiped their finger up or down on the thumbwheel to vary the gain of the bass shelving filter or loudness of the audio in 0.25 dB increments. The goal of the training was to match the bass or loudness level to that of a randomized reference level. Once the listener believed they selected the correct level, they submitted their response and the software informed them of whether they succeeded or not. If they failed to be within the acceptable range (±1 dB), the software informed them of whether their adjustment was too high or too low, and the thumbwheel starting position and reference position re-randomized. If the listener was correct, they moved on to the next trial, with the starting gain of the thumbwheel randomized and the reference gain level randomized. The software recorded how many trials the listener took on each task and the accuracy of their responses.

Fig. 7: Screenshot from bass training section of the software with "infinity thumbwheel" interface.



Following the training tasks, the listeners started the preference level and bass adjustment tasks. The software randomly assigned whether the listener started on the bass or loudness adjustment task first. In the bass task, listeners adjusted a bass shelving filter using an infinity thumbwheel in 0.25 dB increments to their preferred listening level. In the loudness adjustment task, listeners adjusted the loudness level of the songs in 0.25 dB increments to their preferred listening level. In both tasks, the starting position of the thumbwheel was randomized at the beginning of each trial and listeners were give no indication of their position in the range of adjustment. Each of the six tracks was repeated three times throughout the tasks and the playback order of all the trials was randomized.

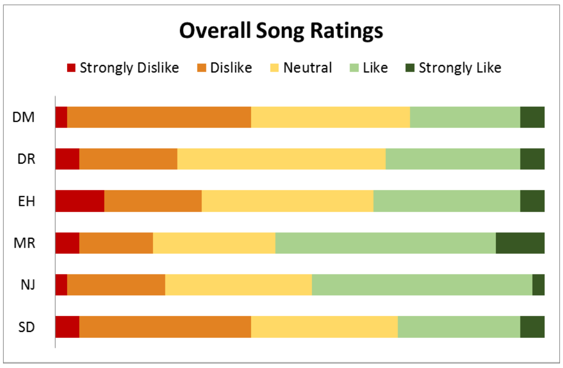
3 Results

3.1 Song Ratings

All six of the tracks were rated as predominantly neutral or better by the overall group of listeners. Of the songs presented, "Uptown Funk"(MR) was rated most favorably while "Black Caffeine"(EH) elicited the most "Strongly Dislike" responses with four total. While listeners had some strong opinions about the Country song in the lineup, "Cousin Dupree"(SD) with 14 "Dislikes" and "Ghosts n Stuff" with 15 "Dislikes" were the most universally disliked tracks tested.

Spearman's rank correlation tests were run to check for a monotonic relationship between song rating and bass or loudness across different groupings. No significant correlations were found.

Fig. 8: Song ratings given by listeners over both headphones and loudspeakers



3.2 Training Data

After analyzing the training data, it was determined that listeners had much more difficulty adjusting bass levels when listening to headphones than over loudspeakers. The mean number of attempts for the bass adjustment task over headphones was 2.73 (SD = 4.25) while it was only 1.44 (SD = 1.12) over loudspeakers. While many listeners had a song they were the least consistent on, song was not a significant factor as verified by a Friedman's test.

Unsurprisingly, the loudness training tests were quite easy for most listeners. Listener performance did not seem to be as strongly affected by playback method as it was for the bass tests. The mean number of attempts for the loudness training tasks over headphones was 1.13 (SD = 0.44) and 1.17 (SD = 0.69) over loudspeakers. In future tests, the range of acceptable responses will likely be reduced to a tighter threshold as even the least experienced listeners had no problem with these tasks.

Performance on the bass training task was a good predictor of listener training level. Of the 12 trained listeners tested all but one completed the bass training tasks with a mean number of attempts less than 2. Some untrained listeners had a great deal of difficulty with the bass training tasks. Since there was no cap on the number of attempts at a particular track, some listeners tried upwards of five times. One listener registered a massive 24 attempts on a single trial! Future tests will be designed to cap the attempts at eight to prevent this level of listener frustration.

3.3 Independent Factor Interaction

Before deciding on statistical testing methods used, the assumptions for parametric tests were checked. It was found that in certain groupings for both the bass and loudness data, the results did not have homogeneity of variances and was not normally distributed. To account for the potential errors that could arise, the data has been analyzed both parametrically and non-parametrically.

A Levene's test was performed to evaluate the homogeneity of variances of the different factor groupings. Bass levels grouped by Playback Method had significantly different variances ($p < 0.05$). The headphone bass results had significantly more variance than the loudspeaker results which points toward the difficulty listeners had adjusting bass consistently over headphones.

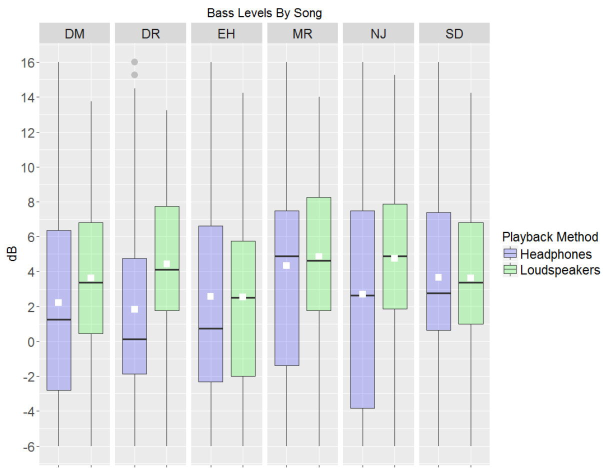
Separate ANOVA tests were performed to analyze interactions of the fixed factors Playback Method (2 levels), Song (6 levels), Listener Training (2 levels), and Listener Hearing Ability (3 levels) for both the bass and loudness tasks. The non-parametric tests used were Friedman's rank sum and Wilcoxon signed rank tests.

For the bass tasks, song was found to be a significant variable ($F(5,80)$, $p = 0.03$). This was confirmed by a Friedman's test. A post hoc Tukey test showed that across playback methods, listeners turned up the bass significantly more on the "Uptown Funk" song than on the other songs.

Playback method was not a significant variable in the ANOVA test for bass, but it proved to be significant non-parametrically. Listeners preferred more bass and higher levels over loudspeakers than over headphones. The bass levels were confirmed to be significantly different between playback methods using a Wilcoxon signed rank test ($Z = -4.54$, $p < 0.001$).

For the loudness tasks, variances of the groupings by playback method were also significant ($p < 0.05$). Playback method was found to be a significant factor ($F(1,16)$, $p = 0.02$) and a Friedman's test verified this result.

Fig. 9: Box plots for the bass levels by song and by playback method. Means are shown as white squares, medians as black lines and outliers as grey circles.



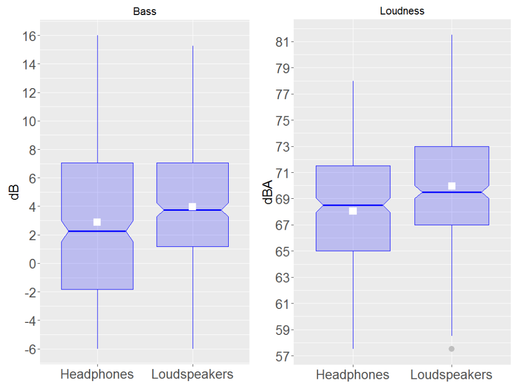
Listener hearing ability did not prove to be a significant factor. This is unsurprising since the majority of listeners with hearing loss had only mild unilateral loss. Listener training was also not a significant factor. That said, trained listeners were much more consistent in their level adjustments and on the training tasks.

3.4 Bass and Loudness Levels

Differences in bass level adjustment varied dramatically across listeners. While individual listeners, particularly trained listeners could generally stay within about a 5 dB window of adjustment consistently, the means of these ranges varied from 14.3 dB for one listener to -3 dB for another. Over headphones, the mean bass gain level was 2.89 dB (SD=6.18) while it was 1 dB higher at 3.9 dB (SD = 4.84) over loudspeakers. The medians were 1.5 dB different: 2.25 dB over headphones and 3.75 dB over loudspeakers.

Loudness level adjustment also varied noticeably across listeners with one listener's mean virtually at full level and another's almost at the minimum loudness. Trained listeners could generally stay within a 3 dB window of adjustment. For the loudness tasks, the mean loudness playback level over speakers was 69.97 dBA (SD = 4.33) while for headphones it was almost 2 dB less at 68.04 dBA (SD = 4.93). The medians were just 1 dB different: 69.5 dBA for loudspeakers and 68.5 dBA for headphones.

Fig. 10: Box plots for the bass and loudness levels adjusted by playback method. Means are shown as white squares, medians as black lines and outliers as grey circles.



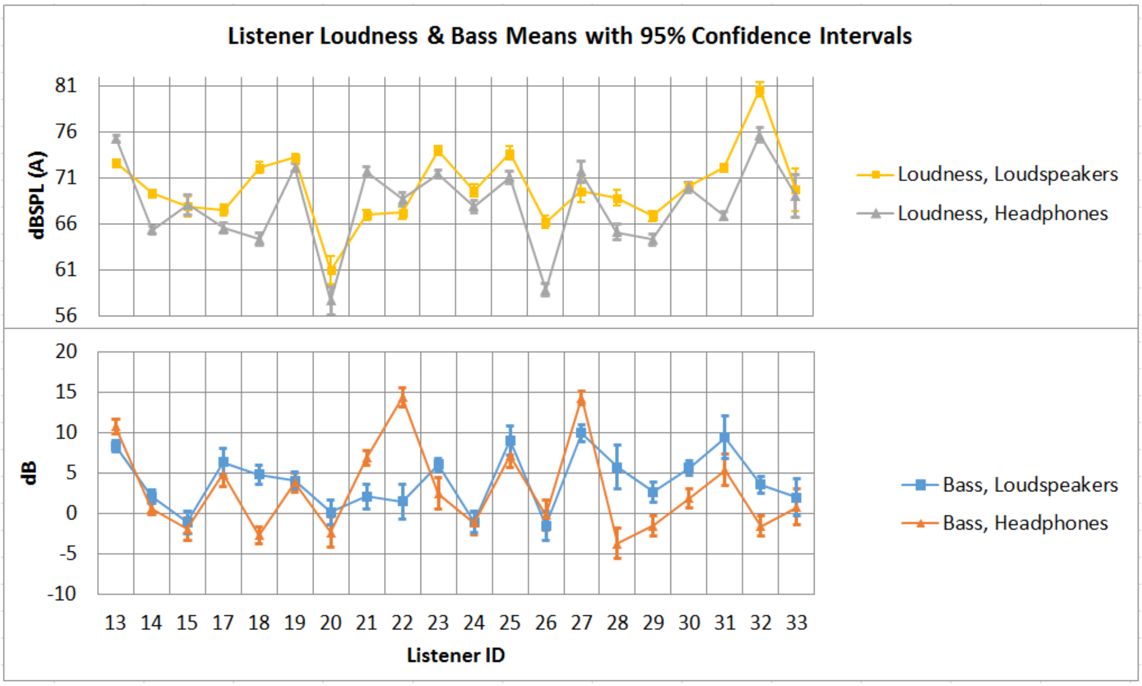
Observing a scatter plot of the data, there appeared to be a strong monotonic relationship between the difference in the mean bass level by listener across playback methods with the difference in mean listener loudness levels across playback methods. A Spearman's rank order correlation was performed and confirmed there was a strong correlation ($r_s = 0.68$, $p = 0.001$) between these differences. This indicates that the direction of bass and loudness changes by listener tended to be the same on either loudspeakers or headphones (i.e. listeners who turned down the bass on headphones also turned down the loudness and vice versa).

4 Discussion

Individual listener preference for both loudness and bass varied considerably in these tests just as in similar studies [1][2][6], which makes generalizing the data challenging. The mean adjusted loudness level of 69.97 dBA over loudspeakers fits neatly into the range between Benjamin's [5] 73.7 dBA mean in automobiles and Dash's [4] 62.5 dBA over televisions. It seems probable that, since the majority of the listeners tested listen critically to soundbars and televisions on a daily basis, their preferred listening levels might be impacted.

When level-matching the headphones and loudspeakers, an additional 2.5 dB boost was added to the headphones to help compensate for the lack of crosstalk and the perceived loss of level that generally occurs between headphones and loudspeakers played back at the same

Fig. 11: Listener Selected Loudness and Bass Levels with 95% Confidence Intervals



level [12]. Considering that the difference in mean levels was about 2 dB in the tests leads one to wonder if this was an effect of the compensation gain or if the listeners preferred listening at lower levels over headphones.

The results of the training and preference adjustment tasks indicate that listeners found adjusting bass levels consistently to be a much more difficult task over headphones than it was over loudspeakers. There are many factors that may have contributed to this increase in difficulty. Since headphone listening is isolated to an in-head experience and excludes the full-body sensations of feeling bass from stereo loudspeakers in a room, listeners have less tactile feedback from which to base their responses. Furthermore, the original mix of the songs was done primarily on stereo loudspeakers, so the intended mix implies the presence of the crosstalk stereo loudspeakers present.

In regards to the bass adjustments, the mean level of 3.93 dB for the bass shelf places the loudspeaker in-room very close to the preferred curve described in [11]. The headphone mean bass level of 2.89 dB is a bit on the low side compared to the results of [2]. This could be due to a number of factors. Most importantly, the program material used in these tests had some very strong low-frequency peaks compared to the songs used in Olive's tests. Also, the "2013 Harman Target curve" was used as the headphone starting point rather than the measurement of loudspeakers in a room at the DRP. The target curve used had considerably less treble and thus would need less bass compensation to sound "balanced."

While listener song rating did not correlate strongly with their individual bass and loudness adjustments, it is interesting to note that the "Uptown Funk" track had significantly higher selected bass levels compared to the other tracks across both headphones and loudspeakers and was the most preferred track across listeners. This may be an effect of the spectral balance of the track which is boosted in the mids and highs above 1 kHz.

After the experiments, the author did brief interviews with the listeners and discovered that listener mood likely impacted some of the responses. For example, listener 22 turned up the bass on average 12.9 dB more over headphones than over loudspeakers. In a post-test interview the listener commented that they were in a good mood during the headphone test and wanted to hear more bass because it was the end of the work day. Another example is listener 20, who turned the bass and loudness down significantly on every test. They expressed that they were aggravated by the music and the test, and were turning the level and bass down as a way to escape. This listener also rated all but one of the songs consistently neutral or lower.

5 Future Work

Several follow-up tests will be run to expand this initial study and address its limitations. First, to reduce the recursive interaction between the tests, in which listeners may have turned up the volume to turn up the bass and vice versa, a test which uses information (e.g. mean bass level) from a listener's first preference test to calibrate the following test will be pursued. Additionally, since the study was quite limited in the scope and size of the listener pool, a much larger study based on a more randomized sample needs to be undertaken. This sample group will need to include a larger group of younger listeners and an equal balance of genders. Finally, the loudspeakers were only tested in a single listening room. For comparison, it would be wise to re-run the loudspeaker portion of the test in different listening environments and specifically in several consumer home environments.

6 Conclusion

This study was designed to evaluate listener preference levels for bass and loudness levels over loudspeakers and headphones. First, listeners evaluated the musical content of all songs used in the tests. Next, they completed several training tasks to familiarize them with the method of adjustment procedure used in the main experiments as well as measure their aptitude. Following the training tasks, listeners completed two eighteen-trial tests in which they adjusted the bass and loudness level in 0.25 dB increments to their preferred setting using a virtual infinity knob, which had detents but no reference or level markings. The results are summarized as follows:

1. The mean bass gain level was 2.89 dB (SD = 6.18) over headphones and 1 dB higher at 3.9 dB (SD = 4.84) over loudspeakers.
2. The mean loudness level was 68.04 dBA (SD = 4.93) over headphones while over loudspeakers it was 2 dB more at 69.97 dBA (SD = 4.33)
3. Listeners had much more difficulty adjusting the bass levels consistently over headphones than loudspeakers.
4. The variance of both the bass and loudness preference results was significantly higher over headphones than over loudspeakers.
5. Song had a significant effect on the results of the bass preference tests. Listeners tended to increase the bass more on the "Uptown Funk" track.
6. Song rating was not directly correlated with loudness or bass levels selected.
7. Listener training did not have a significant impact on their selected bass or loudness range. Trained listeners were more consistent in their range of selection than untrained listeners.
8. There was a strong correlation between the difference in the mean bass levels by listener across playback methods with the difference in mean listener loudness levels across playback methods. In other words, listeners who turned up the bass over headphones or loudspeakers usually turned up the loudness as well.

Acknowledgements

Samsung Research America fully supported this work. Thanks to all the participants of the Audio Lab Listening Team. Special thanks to Victoria Suha, Adrian Celestinos, Glenn Kubota, and Eric Klerks.

References

[1] Olive, S., Welti, T., and McMullin, E., "Listener Preferences for Different Headphone Target Response Curves," *presented at the 134th Convention, Audio Eng., Soc., preprint 8867*, 2013 May.

[2] Olive, S. and Welti, T., "Factors that Influence Listeners' Preferred Bass and Treble Balance in Headphones," *presented at the139th AES Convention, Audio Eng., Soc., preprint 9382*, 2015 October.

[3] Olive, S., Welti, T., and Khonsaripour, O., "The Preferred Low Frequency Response of In-Ear Headphones," *presented at the AES International Conference on Headphone Technology*, 2016 August.

[4] Dash, I., Power, T., and Cabrera, D., "The relation between preferred TV programme loudness, screen size and display format," *presented at the 134th AES Convention, Audio Eng Soc., preprint 8817*, 2013 May.

[5] Benjamin, E. and Crockett, B., "Preferred Listening Levels in the Automotive Environment," *presented at the 119th Convention, Audio Eng. Soc., preprint 6533*, 2005 October.

[6] Worthington, D., Siegel, J., Wilber, L., Faber, B., Dunckley, K., Garstecki, D., and Dhar, S., "Comparing two methods to measure preferred listening levels of personal listening devices," *Journal of the Acoustical Society of America*, 125, pp. 3733–3741, 2009.

[7] King, R., Leonard, B., and Sikora, G., "The Effects of Monitoring Systems of Balance Preference: A comparative study of mixing on headphones versus loudspeakers," *presented at the 131 AES Convention, Audio Eng Soc., preprint 8566*, 2011 October.

[8] McMullin, E., Celestinos, A., and Devantier, A., "Environments for Evaluation: The Development of Two New Rooms for Subjective Evaluation," *presented at the 139th AES Convention, Audio Eng. Soc., preprint 9460*, 2015 October.

[9] 1116-1, I. R. B., "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multi-channel Sound Systems," *International Telecommunications Union Radiocommunication Assembly*, 1997.

[10] Devantier, A., "Characterizing the Amplitude Response of Loudspeaker Systems," *presented at the 113th Audio Eng. Soc. Convention*, 2002 October.

[11] Olive, S., Jackson, J., Devantier, A., Hunt, D., and Hess, S., "The Subjective and Objective Evaluation of Room Correction Products," *presented at the 127th AES Convention, Audio Eng Soc, preprint 7960*, 2009 October.

[12] Bech, S. and Zacharov, N., *Perceptual Audio Evaluation*, John Wiley & Sons, Ltd, 2006.

[13] Lin, F., Niparko, J., and Ferrucci, L., "Hearing Loss Prevalence in the United States," *Arch Intern Med.*., 171(20), p. 1851–1852, 2011 Nov 14.