Sound Quality Measurements in Headphones

P.H.W. Leong, Y.S. Moon, W.K. Sim
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, NT Hong Kong

D.W.P. Lam
Globe Audio Products Mfty. Ltd.
No 5 & 6, 5/F 43-47 Shan Mei St, World Wide Industrial Centre
Fo Tan, Shatin, NT Hong Kong

Abstract

A technique for identifying important measurable headphone parameters is presented. Fifty two headphones of the same model, were computer clustered into 4 groups based on frequency response and listening tests performed to determine subjective sound quality. Correlation at the 95% significance level with the voice coil impedance was found but no correlation with distortion or frequency measures was observed.

List of Important Symbols

A_{46Hz}	Gain at 46 Hz
A_{1kHz}	Gain at 1 kHz
$A_{11547Hz}$	Gain at 11547 Hz
D_2	Second harmonic distortion (1 kHz input signal)
D_3	Third harmonic distortion (1 kHz input signal)
R_E	DC resistance of the voice coil
F_s	Resonant frequency of driver
Q_{es}	Total Q of driver at Fs considering only electrical resistances
Q_{ms}	Total Q of driver at Fs considering only non–electrical resistances
Q_{ts}	Total Q of driver at Fs considering all system resistances
L_{1k}	Inductance of voice coil at 1 kHz
L_{10k}	Inductance of voice coil at 10 kHz

0 Introduction

The "sound quality" of an audio system is a very subjective measure which is not easily characterised by quantitative measurements. Although there is great interest in high fidelity sound reproduction, surprisingly little research has been conducted into finding correlations between subjective listening tests and measurable parameters. By far the most common quantitative measure of headphone performance is frequency response and experiments were conducted to test its ability to discriminate between low fidelity mass produced headphones. Other common quantitative measures of headphone fidelity include various measurements of distortion and the Thiele–Small parameters [1] of the voice coil.

This paper presents an approach to testing headphones which comprises of the following steps features

1. in order to reduce the number of headphones to undergo subjective listening tests, a large number of headphones are clustered based on frequency response into a small number of clusters and a small number of headphones are selected from the clusters

2. listening tests are conducted with each headphone being tested several times and compared against several different headphones by different people

3. in order to simplify the task required by the listener, each is asked to grade 4 randomly selected headphones by making three binary comparisons
4. all the measurable parameters of the headphones tested are measured
5. the correlation coefficient between measured parameters and listening test results is computed to provide a single figure measure of the parameter's ability to predict the listening test result.

This approach can be used to identify those parameters which directly affect the sound quality of the headphone (if any exist), making it possible to make design and/or manufacturing changes which lead to an overall improvement in the sound quality of the product.

1 Experimental Design

1.1 Frequency Response Clustering

Fifty two mass produced headphones of the type typically supplied with portable tape recorders and CD players were studied. In order to minimise the manufacturing variations, all of the headphones were of the same model produced on the same production line on the same day. The frequency response for each headphone was measured using Audiomatica Srl's CLIO system [2]. A plot of the frequency response of all headphones is shown in Figure 1.

It can be seen from the frequency response measurements of Figure 1 that the low frequency gain varies dramatically between headphones. We believe this was due to the mechanical device used to hold the headphones and the low frequency measurements should not be regarded as being accurate.

A clustering technique was applied to group headphones with similar frequency response. The procedure proceeds as follows

1. principle component analysis (PCA) was applied to the magnitude of the frequency response to extract the salient features from the data and achieve dimensionality reduction [3]
2. the reduced dimension data was then clustered using the K-means algorithm [4] to group the data into clusters.

The above clustering technique was applied to cluster the frequency responses of the 52 different headphones into 4 clusters. The result of this clustering process is shown in Figure 2, each subplot showing the frequency response of all headphones grouped in a particular cluster. The frequency distribution of the clustering results are shown in Table 1.

If frequency response were a good measure of headphone quality, one might expect that a single cluster would contain a disproportionally high number of similar sounding

Cluster	Frequency
1	23
2	19
3	8
4	2

Table 1: Frequency distribution of headphone clusters.

headphones. Since listening tests based on 52 different headphones would be too large an undertaking, a subset of these headphones based on the clustering results was selected.

Random samples from each cluster were taken with the number dependent on the number of headphones in the cluster. Since cluster 1 had the most headphones, three samples were taken from it. Two headphones were taken from clusters 2 and 3 and a single sample from cluster 4. This subset of 8 headphones was chosen as a representative set of the different characteristics of the original 52 headphones to be used in the listening tests.

A sample size of 8 headphones was selected since it made the number of listening tests small enough to be feasibly conducted in our laboratory using a simple randomly drawn listening test. A much larger sample size would increase our confidence in the results but would require an enormous number of listening tests. It is also possible to perhaps use a smaller number of listeners and ask them to assign a numerical score to each headphone. This procedure would enable a much larger number of headphones to be tested, however, listener fatigue and personal preference may lead to unreliable results. It must be remembered that the goal of the headphone industry is to produce headphones that sound good to the majority of listeners rather than satisfy a small number of critical headphone reviewers.

1.2 Additional Headphone Measurements

Additional parameters of the 8 selected headphones were also measured using Audiomatica Srl's CLIO system. The first set of parameters, A_{46Hz}, A_{1kHz}, $A_{11547Hz}$ were taken from the frequency response and correspond to the nominal low, midrange and high frequency responses of the headphones. The second set of measurements were D_2 and D_3 which correspond to the distortion levels of the headphones to a 1 kHz sinusoidal input signal. Finally, the Thiele–Small parameters [1] R_E, F_s, Q_{es}, Q_{ms}, Q_{ts}, L_{1k} and L_{10k} were measured which relate directly to the voice coil.

2 Listening Tests

The system used to evaluate the headphones was typical of the type of system with which such headphones would be used. It comprised of a Sony D-465 Discman CD player connected to a distribution amplifier. The amplifier, constructed from Burr-Brown OPA604 opamps with a gain of 2, was used so that the 4 headphones could be evaluated simultaneously without disconnecting headphones. The song "Desafinado" from the CD "Jazz Samba" (Verve records 810 061-2) was used for all of the evaluation tests. This piece was selected for its transient nature, being a mix of drums, tenor sax, bass and guitar.

Using the clustered set of 8 headphones, a simple informal listening test was conducted in the following manner: a computer program was used to randomly draw four headphones in a random order. Each subject was asked to choose their preference between the first and second headphones and then the third and fourth headphones on the list. Then the subject was asked to select between the two winning headphones which determined the headphone they most preferred. In this manner, a winning headphone was selected from 4 different headphones based on three comparisons.

3 Results

3.1 Listening Tests

A total of 51 different people were used as subjects and the frequency distribution of the results is shown in Table 2. A single figure of merit was derived from the listening results by assigning a score of 3 to each occurrence of a first place, 1 for a second place and 0 for a third place result and then dividing by the total number of results to normalise the data. This value is shown in the "Score" column of Table 2.

Looking at the scores associated with each headphone, they are grouped around three values. G was clearly superior to the other headphones, having obtained more 1st place results and a higher score than the others. Headphones C, D, A and B followed although their differences in score were not large. The worst scoring (and presumably the worst sounding) headphones were F, H and E.

The four most favoured headphones according to score (G, C, D and A) originated from four different clusters. This indicates that a strong relationship between frequency response and listening test preference does not exist.

3.2 Correlation with Measured Parameters

Measurable parameters from headphones are often optimised by designers in the hope that improvement in such a parameter will directly relate to the sound quality of a headphone. Common design criteria including wide frequency response and low distortion. Such an assertion can only be true if the parameter is strongly correlated with a listener's perceived sound quality of that headphone. The parameters described in Section 1.2 were all tested for their correlation to the score obtained from listening tests.

From elementary statistics (e.g. [5]), the t test for testing correlation coefficients is

#	Cluster	1st place	2nd place	3rd place	Score
A	1	8	4	14	1.0769
B	4	5	9	10	1.0000
C	3	7	7	11	1.1200
D	2	7	4	12	1.0870
E	4	7	5	20	0.8125
F	2	3	7	9	0.8421
G	1	10	8	13	1.2258
H	1	4	7	12	0.8261

Table 2: Frequency distribution of listening test results. The first column is the headphone identification letter, the second column is the cluster number from which it was selected, the 3rd–5th columns represent the frequency of a 1st–3rd ranking respectively and the last column represents the normalised score given to that headphone.

given by

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \tag{1}$$

where n is the number of paired observations and r is the correlation coefficient. From the value of t thus computed, a level of significance can be found using a table which is included in most statistics textbooks (e.g. [5]). For the experiments conducted, $n = 8$ and if a 95% level of significance two–tailed test is desired, $r > 0.632$ for the null hypothesis to be rejected (i.e. there is significant correlation between the parameter and the listening test).

3.2.1 Frequency Parameters

Table 3 shows the gain of each of the 8 headphones at low, midrange and high frequencies. It can be seen that none of the correlation values, r, are greater than 0.632. Thus testing the correlation coefficients for significance at the 95% level, we cannot reject the null hypothesis and thus cannot find significant correlation between these parameters and the score. This confirms the findings in Section 3.1 that frequency response is not a good measure for determining the sound quality of these headphones.

3.2.2 Distortion Parameters

Distortion is a commonly used measure of the nonlinearities of a system. The measurements shown in Table 4 show measurements of the 2nd and 3rd harmonic distortions[1]

---

[1] The total harmonic distortion (THD) was not obtained because our version of the CLIO software does not support this measurement.

#	Score	A_{46Hz} (left, dB)	A_{46Hz} (right, dB)	A_{1kHz} (left, dB)	A_{1kHz} (right, dB)	$A_{11547Hz}$ (left, dB)	$A_{11547Hz}$ (right, dB)
G	1.23	80.23	77.67	85.71	85.71	68.14	76.09
C	1.12	85.1	84.92	88.02	88.42	79.63	75.57
D	1.09	78.41	81.28	85.49	86.18	69.83	75.2
A	1.08	84.63	84.63	87.51	87.51	80.02	80.02
B	1.00	84.84	86.48	88.72	88.42	81.25	83.61
F	0.84	82.42	84.15	84.77	87.69	77.93	79.6
H	0.83	84.27	82.21	87.69	87.15	76.37	70.82
E	0.81	77.29	82	86.68	85.32	75.27	71.98
r		0.0195	0.0212	0.0843	-0.286	-0.344	0.280

Table 3: Measured frequency response parameters from the 8 headphones (sorted by score). The bottom row is the correlation (r) of the parameter with the listening test score.

for the 8 headphones measured. As with the frequency parameters, none of these measurements were correlated with the listening test results.

3.2.3 Thiele–Small Parameters

The Thiele–Small parameters are related only to the voice coil of the headphone. The values measured in Table 5 do not show good correlation with the listening test values. However, all of the measurements of Table 6 have $r > 0.632$ and hence show a statistically significant correlation (at the 95% level) with the listening test score value.

These parameters are all related to the impedance of the voice coil. Production of these headphones is done in a very manual fashion and variations in the coil wire length and inductance can result from the manner in which the components are placed in the coil winding machine. Work is presently underway to find more headphones from the same batch which have R_E, L_{1K} and L_{10K} parameters similar to those of headphone G and conduct further listening tests to determine whether for this particular model of headphone, the sound quality is directly related to these parameters.

4 Conclusion

A method was presented for identifying correlations between subjective listening tests and measurable parameters. Fifty two same model headphones produced from the same production line were clustered into 4 different groups based on frequency response. From these clusters, eight headphones were selected for listening tests. The listening tests involved 51 different listeners who were asked to assess four randomly drawn headphones based on three binary decisions. A score related to the subjective sound quality of the

#	Score	D_2 (left, dB)	D_2 (right, dB)	D_3 (left, dB)	D_3 (right, dB)
G	1.23	49.68	33.24	29.81	20.72
C	1.12	41.45	41.47	25.36	27.05
D	1.09	30.19	31.08	24.76	31.54
A	1.08	37.25	37.25	28.52	28.52
B	1.00	39.76	42.17	26.12	22.97
F	0.84	38.87	40.31	25.77	10.13
H	0.83	39.21	40.46	33.64	23.6
E	0.81	28.87	15.1	23.95	25.31
r		0.528	0.219	-0.008	0.376

Table 4: Measured distortion parameters from the 8 headphones (sorted by score). The bottom row is the correlation (r) of the parameter with the listening test score.

#	Score	Q_{ms} (left)	Q_{ms} (right)	Q_{es} (left)	Q_{es} (right)	Q_{ts} (left)	Q_{ts} (right)	F_S (left, Hz)	F_S (right, Hz)
G	1.23	0.54	0.48	9.04	8.09	0.51	0.45	128.72	140.33
C	1.12	0.65	0.58	8.59	7.79	0.6	0.54	144.14	133.65
D	1.09	0.57	0.64	8.81	7.7	0.53	0.59	139.12	132.48
A	1.08	0.6	0.53	8.2	7.31	0.56	0.49	136.34	129.85
B	1.00	1.27	0.67	9.36	7.31	1.12	0.61	139.95	140.33
F	0.84	0.67	0.53	11.06	6.98	0.63	0.5	141.56	127.3
H	0.83	0.53	0.58	9.08	8.51	0.5	0.54	134.85	141.56
E	0.81	0.66	0.42	8.72	7.83	0.61	0.4	139.12	131.34
r		-0.0739	0.185	-0.430	0.0325	-0.0834	0.164	-0.315	0.201

Table 5: Measured Thiele–Small parameters (I) from the 8 headphones (sorted by score). The bottom row is the correlation (r) of the parameter with the listening test score.

#	Score	R_E (left, Ω)	R_E (right, Ω)	L_{1K} (left, mH)	L_{1K} (right, mH)	L_{10K} (left, mH)	L_{10K} (right, mH)
G	1.23	34.7	35.4	1.09	1.17	0.16	0.17
C	1.12	34.8	34.3	1.08	1.14	0.16	0.16
D	1.09	34.9	34.3	1.09	1.11	0.16	0.16
A	1.08	35	34.3	1.1	1.13	0.16	0.16
B	1.00	34.3	34.2	1.05	1.17	0.16	0.16
F	0.84	33.7	33	0.99	1.1	0.15	0.15
H	0.83	34.5	32.5	0.92	1.01	0.15	0.14
E	0.81	34.4	33.2	1.03	1.09	0.15	0.15
r		0.660	0.955	0.822	0.735	0.916	0.918

Table 6: Measured Thiele–Small parameters (II) from the 8 headphones (sorted by score). The bottom row is the correlation (r) of the parameter with the listening test score.

headphone was determined from these tests and the correlation with various measurable frequency, distortion and Thiele–Small parameters determined.

Using the t test for testing correlation coefficients for significance, no correlation with frequency parameters was found. However, correlation at the 95% significance level with parameters related to the impedance of the voice coil were identified. The approach presented allows for parameters which directly affect the subjective sound quality of the headphone to be identified, making it possible to make design and/or manufacturing changes which lead to an overall improvement in the sound quality of a product.

References

[1] R. Small, "Vented-box loudspeaker systems part ii: Large-signal analysis," *J. Audio Eng. Soc.*, vol. 21, pp. 438–444, July/August 1973.

[2] A. S.R.L., *CLIO User's manual.* 1998.

[3] I. Jolliffe, *Principle Component Analysis.* New York: Springer–Verlag, 1989.

[4] J. Hartigan, *Clustering Algorithms.* New York: Wiley, 1975.

[5] J. Williams, *Statistical Methods.* University Press of America, 1996.

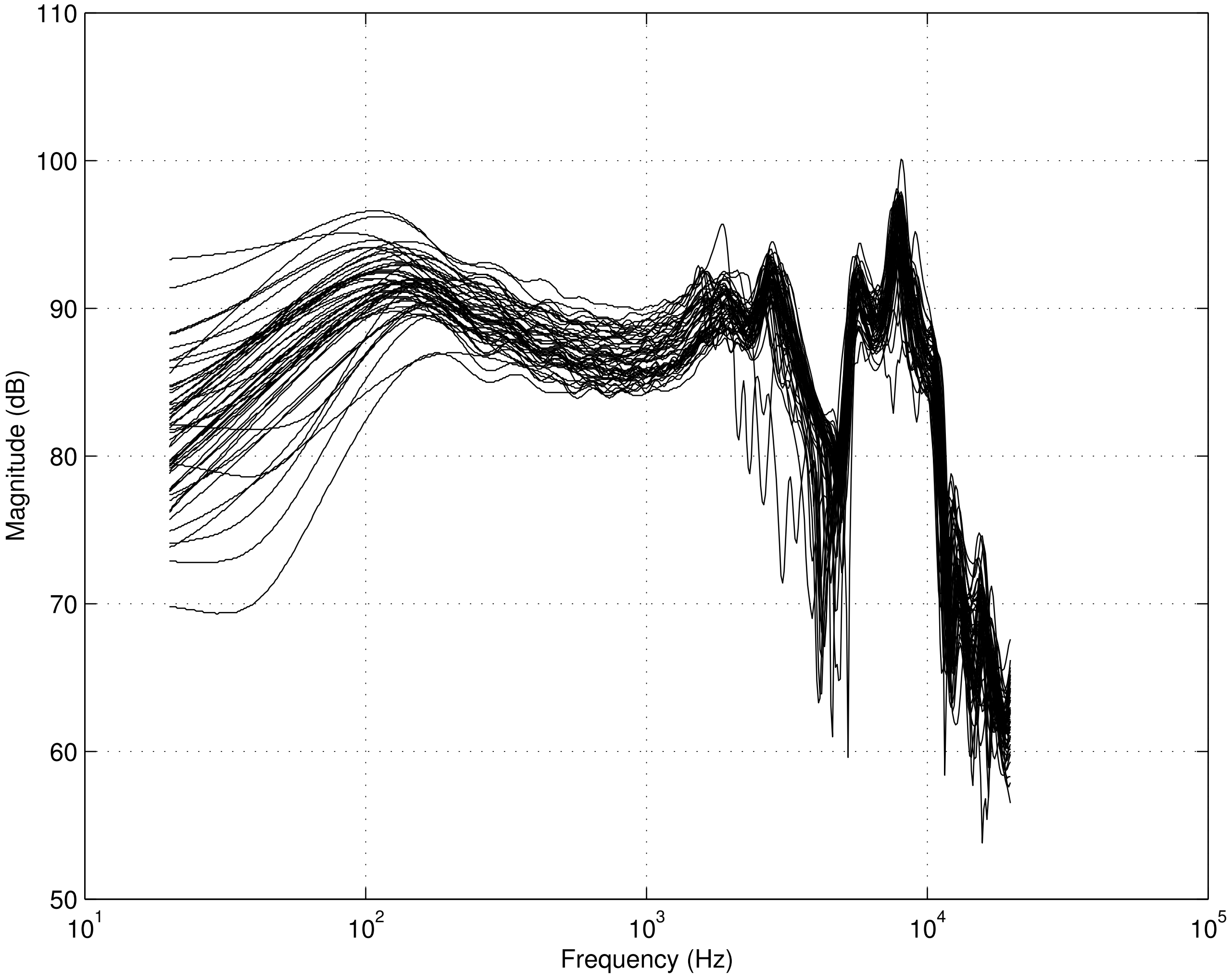

Figure 1: Frequency response of all headphones.

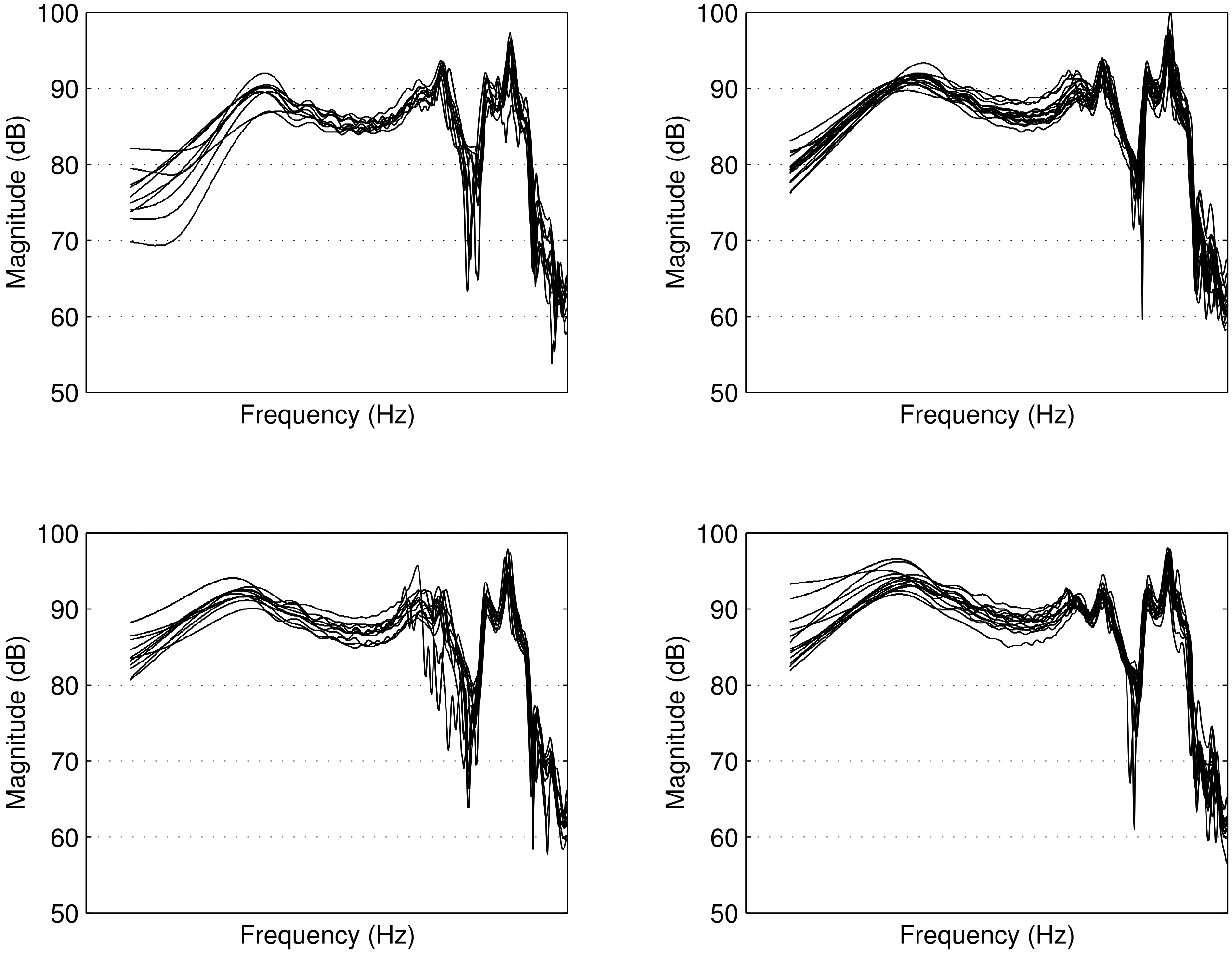

Figure 2: Frequency response of all headphones grouped by cluster.

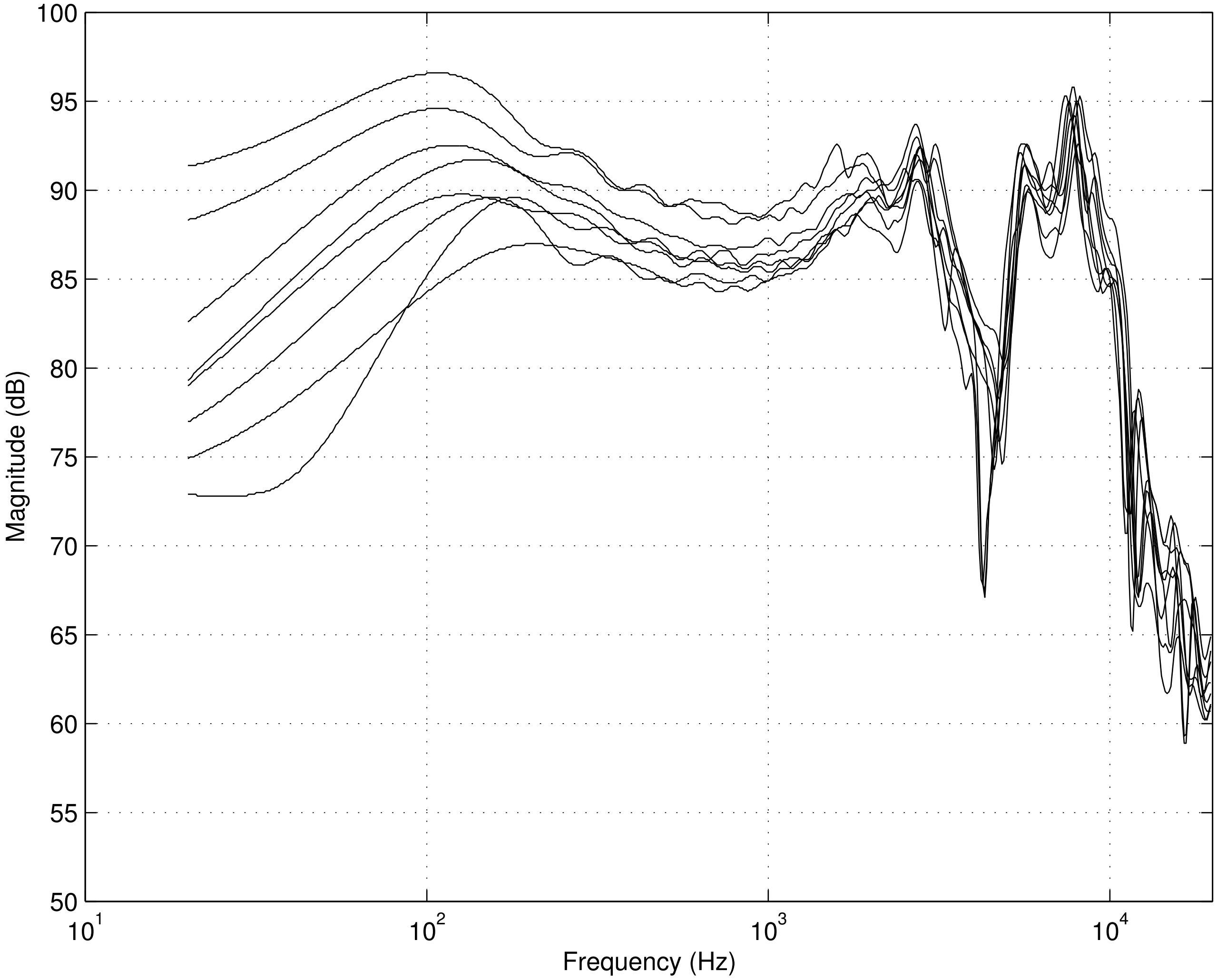

Figure 3: Frequency response of the selected headphones.