Apple Headset (AHS) calibration

Frans Coninx, Andrzej Zarowski, Nicolas Wack, Erwin Offeciers

(last edited: 2015-10-31)

Introduction – Apple Headset (AHS)

In September 2012, Apple introduced a redesign of the earphones they had been shipping by default with their mobile devices, while at the same time rebranding them to "Apple EarPods". The new design was lauded by many, for their aesthetic qualities but also with respect to sound quality. Of particular relevance we can note that these are the default earphones bundled with all new iPhones and iPods produced by Apple, and as such gathering data on these earphones is of prime importance as it is possible to safely assume that an app running on those devices will have access to these earphones.

The EarPods Apple Headset (AHS) will be used by end-users for self-assessment of tone detection thresholds on iPod/iPhone hardware devices while using the DuoTone software.

The present study was conducted in order to collect calibration data for the AHS, based on an intra-subject comparison of tone detection threshold data, collected with a calibrated system as well as with the AHS.

For this purpose, thresholds were measured in a group of normal hearing adult subjects using two different procedures: one method (test A) used the calibrated clinical audiometry equipment with results in dBHL, the other method (test B) used the not yet calibrated AHS. By combining the individually determined thresholds with the tests A and B, the calibration of the AHS could be performed.
This can be described as crowd-calibration and the details of the procedure can be found in the provisional US patent 62/190,334. A similar method can be found in [Larrosa, 2015][1].

---

[1] Larrosa F, Rama-Lopez J, Benitez J et al. Development and evaluation of an audiology app for iPhone/iPad mobile devices. Acta Otolaryngol, 2015 Jul 4

Equipment and Methods

Pure tone detection thresholds were measured in two test conditions (A and B). In both test conditions the same DuoTone measurement procedure for the detection threshold of pure tones was applied (see also the report “Antwerp Study DuoTone Validation” from October 2015).

In test A the iPod was connected to a calibrated combination of an Interacoustics Equinox audiometer and a Sennheiser HDA200 headphone.
In test B the iPod was connected to an Apple Headset (AHS).

DuoTone thresholds were measured for each subject on the ear of their choice (sometimes both) at 12 audiometric frequencies, presented in 6 frequency pairs: (125;1k5), (250;2k), (500;4k), (750;6k), (1k;8k), (3k;12k) [Hz].
The presentation order of these DuoTone frequency pairs was randomized.

THR-A being the threshold obtained with test A equals AA(f,i) + HDA200cal(f), with

- AA(f,i) being the tone threshold in test A at frequency /f/ for the subject /i/, in dBFS
- HDA200cal(f) being the calibration value for the HDA200 headphones at frequency f in dBHL.

THR-B being the threshold obtained with test B equals BB(f,i) + AHS(f), with

- BB(f,i) being the tone threshold in test B at frequency /f/ for the subject /i/, in dBFS
- AHS(f) being the calibration value for the AHS at frequency f.

THR-A and THR-B are supposed to be equal, except from intra-subject threshold differences. Therefore: $BB(f,i) + AHS(f) = AA(f,i) + HDA200cal(f) + error(f, i)$. This difference though can be cancelled by averaging over multiple instances of the measurements, which yields the following:
The AHS calibration can be calculated as $AHS(f) = \mathbb{E}[AA(f,i) - BB(f,i)] + HDA200cal(f)$.

Calibration of test A

The Interacoustics Equinox audiometer was calibrated following regular clinical procedures.

The DuoTone procedure on the iPod device uses an internal dB scale, implemented in the software. Without any software attenuation, the maximum output is 0 dBFS (dB Full Scale). For test condition 1, each of the 12 test frequencies on the iPod was calibrated, at the maximum available level on the iPod, being 0 dBFS. The iPod was connected to the audio-input of the Equinox audiometer and the output level on the Equinox audiometer set to 80 dBHL. Table 1 summarizes the calibration results for the combination of iPod signal source, Equinox audiometer and HDA200 headphones in dB SPL as well as in dB HL.

	Frequency (Hz)											
	125	250	500	750	1000	1500	2000	3000	4000	6000	8000	12000
in dB SPL	85,4	86,8	87,5	88,4	90,4	92,0	90,1	88,7	88,8	89,2	87,9	84,0
in dB HL	54,9	68,8	76,5	82,4	84,9	86,5	85,6	86,2	79,3	72,2	70,4	56,0

Table 1: HDA200cal (f) values in dBSPL and dBHL at iPod output signal 0 dBFS and equinox input sensitivity set at 80 dBHL

Subjects

All subjects were team members of the ENT/audiology department of the St.Augustinus clinic in Antwerp. Each subject was informed about and invited to participate in the tests.
During one day, 24 subjects decided to participate. Twenty of them were tested only on one single ear, whereas 4 subjects decided to test two ears with a time break in between. In total N=28 ears could be tested. Out of the 24 subjects, 13 were female and 11 were female. The youngest subject was 20 years old, and the oldest 65, with an average age over all the subjects of 42.2 years.

Results

Table 2 shows the AHScal(f,i) data for each of the 12 frequencies (f) and 28 ears (i). These values were calculated using formula AHS(f) = AA(f,i) - BB(f,i) + HDA200cal(f). See above under “equipment and methods”.

A statistical analysis of these data is shown in Table 3.
The average value is defined to be the AHScal(f) value.

One ear (number 7 in table 2) causing outlier values, for unknown reasons, was excluded from the data set.
The standard deviation of the AHScal values ranges from 4,8dB (at 4kHz) to 7,8dB (at 125Hz and 12000Hz). The standard deviation averaged over the 12 frequencies is 6.2dB.

Finally the median values were calculated.
The difference between average and median values is very low. Averaged over the 12 frequencies the difference is only 0,2 dB. The largest difference is 1.6 dB (at 2 kHz).
In summary, the average and median values are summarized in Figure 1.
For each frequency the average values (plus and minus one standard deviation) and the median values are plotted.

	Frequency (Hz)											
	125	250	500	750	1000	1500	2000	3000	4000	6000	8000	12000
1	86,5	98,8	99,8	95,8	79,9	103,2	103,9	106,2	99,3	105,6	102,0	89,3
2	94,9	105,4	88,1	90,8	88,2	101,5	103,9	106,2	102,6	112,2	98,7	86,0
3	86,5	90,4	86,5	94,1	116,5	106,5	112,2	107,9	105,9	102,2	107,0	91,0
4	82,4	97,1	101,5	100,8	98,2	98,2	105,6	101,2	104,3	102,2	98,7	89,3
5	86,5	93,8	101,5	100,8	101,5	104,8	108,9	111,2	115,9	102,2	97,0	81,0
6	88,2	95,4	99,8	110,8	109,9	104,8	102,2	111,2	104,3	105,6	100,4	82,7
7	88,2	85,4	93,1	89,1	96,5	139,8	135,6	99,6	105,9	127,2	90,4	74,3
8	74,9	92,1	106,5	102,4	106,5	101,5	103,9	104,5	102,6	107,2	102,0	91,0
9	89,9	85,4	96,5	104,1	98,2	104,8	97,2	111,2	109,3	103,9	93,7	84,3
10	78,2	90,4	88,1	94,1	94,9	88,2	110,6	117,9	110,9	112,2	98,7	86,0
11	86,5	92,1	99,8	104,1	98,2	103,2	100,6	119,6	109,3	110,6	102,0	91,0
12	88,2	90,4	96,5	100,8	104,9	111,5	103,9	111,2	102,6	110,6	103,7	91,0
13	71,5	93,8	93,1	95,8	98,2	108,2	105,6	102,9	112,6	107,2	95,4	61,0
14	79,9	83,8	96,5	99,1	99,9	103,2	112,2	106,2	99,3	100,6	102,0	81,0
15	81,5	90,4	88,1	89,1	96,5	104,8	123,9	97,9	107,6	88,9	82,0	82,7
16	81,5	93,8	89,8	99,1	96,5	104,8	108,9	97,9	112,6	107,2	107,0	89,3
17	103,2	98,8	106,5	99,1	98,2	103,2	113,9	101,2	100,9	103,9	100,4	84,3
18	99,9	95,4	91,5	97,4	99,9	109,8	108,9	116,2	102,6	108,9	100,4	89,3
19	88,2	97,1	96,5	102,4	103,2	106,5	107,2	107,9	104,3	102,2	98,7	86,0
20	84,9	105,4	99,8	107,4	108,2	106,5	115,6	111,2	107,6	113,9	92,0	82,7
21	98,2	100,4	93,1	115,8	108,2	106,5	105,6	87,9	97,6	102,2	100,4	92,7
22	91,5	95,4	101,5	110,8	88,2	104,8	105,6	99,6	107,6	105,6	98,7	83,5
23	98,2	102,1	108,1	110,8	106,5	106,5	102,2	111,2	104,3	110,6	102,0	84,3
24	94,9	98,8	103,1	102,4	104,9	108,2	98,9	109,6	100,9	103,9	100,4	82,7
25	86,5	92,1	93,1	100,8	106,5	113,2	117,2	106,2	109,3	97,2	97,0	66,8
26	81,5	88,8	91,5	97,4	101,5	84,8	103,9	99,6	100,9	102,2	97,0	75,2
27	86,5	92,1	96,5	97,4	89,9	108,2	105,6	114,6	112,6	107,2	105,4	73,5
28	76,5	87,1	94,8	102,4	98,2	103,2	105,6	106,2	107,6	103,9	98,7	71,8

	Frequency (Hz)											
	125	250	500	750	1000	1500	2000	3000	4000	6000	8000	12000
Average = AHScal(f)	86,9	94,3	96,6	100,9	100,0	104,1	107,2	106,8	105,8	105,2	99,3	83,3
Stand.deviation	7,8	5,5	6,0	6,3	7,7	6,0	5,9	7,0	4,8	5,2	4,9	7,8
median	86,5	93,8	96,5	100,8	99,0	104,8	105,6	106,2	105,1	105,6	99,5	84,3

Table 3: Averaged data (N=24 ears) For A

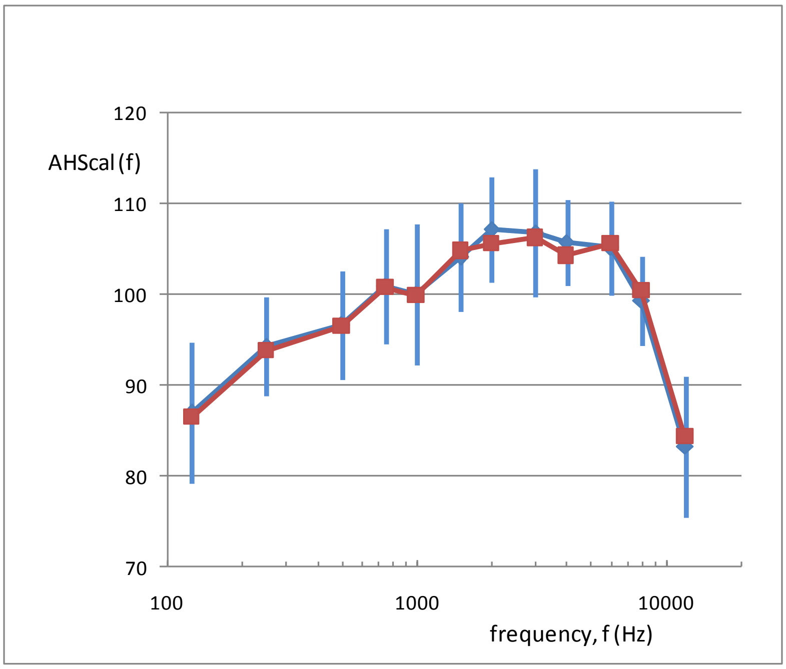

Figure 1: Average and standard deviation (in blue) and median values (in red) for the AHScal(f).

Discussion

The frequency dependent AHScal values could successfully be obtained and can now be used for calibrated pure tone threshold measurements using the combination of iPod and AHS. For any other combination, calibration values have to be collected once again.

The data collected in this subjective calibration procedure has proven to be very stable and reliable.
A mean standard deviation of 6.2 dB may be interpreted as very low. Several factors will have contributed to spread of data:

- Intra-individual differences between the subjective measurement of the threshold with the HDA200, causing a standard deviation of 5 dB (order of magnitude)
- Intra-individual differences between the subjective measurement of the threshold with the AHS, also causing a standard deviation of 5 dB (order of magnitude)
- Variation in the insertion/positioning of the AHS in the external ear canal, causing variations of unknown magnitude (See follow-up study included hereafter that discusses the extent of this factor)

These variations might add up to a spread of 10dB.
Therefore, the mean standard deviation of only 6.2 dB in this study is showing high stability of the calibration data.

Reliability of the previous measurements

In order to ensure that the values measured in the previous experiment are reliable enough to be depended upon, we need to characterize what factors can influence the measured output level, and by how much. There are 2 such factors that we have identified:

- The difference in the way the AHS are inserted in the ear (insertion / re-insertion spread)
- The difference in output between different instances of the same EarPods model (Intra-headset variability)

Insertion / re-insertion spread

As the exact sound level at the eardrum level might vary depending on the way the earphones have been positioned in the ear, a follow-up study has been performed where special care has been taken to try to reproduce as much variability as possible for a given AHS by trying different way of positioning it in the ear (insertion/reinsertion) and comparing the resulting values.

This study has been performed by setting up a GRAS 45CB acoustic testing fixture with an artificial ear. The same iPod has been used in all cases and was producing the same pink noise test signal. Absolute SPL values at the eardrum level were then measured for each 1/3-octave band from 125Hz up to 20KHz, for 16 insertions of the AHS in the artificial ear. The standard deviation (representing the average error range) can be seen in figure 2.

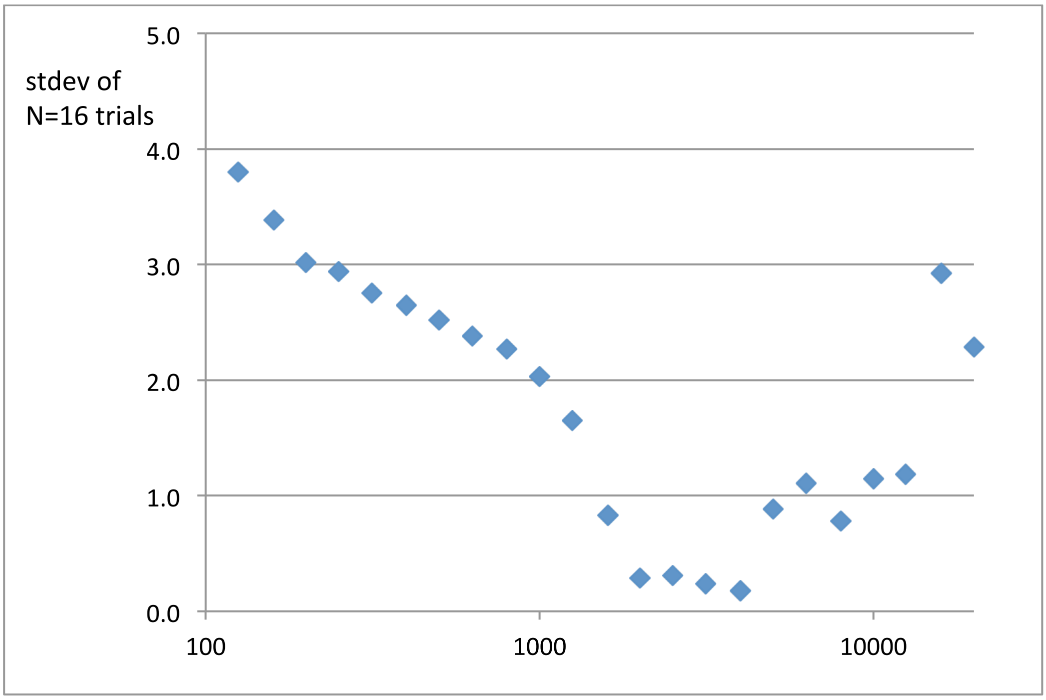

Figure 2: standard deviation of the output level of AHS after multiple insertions/reinsertions

From this, we can see that the positioning of the AHS in the ear has little effect on the sound output level produced by them as recorded at the eardrum level. This is especially true for the frequency region between 1KHz and 10KHz, where the standard deviation is close to, or lower than 1dB.

It is to be noted also that the insertions done in this experiment have been done by inserting the AHS in different positions (as long as they did not fall from the ear), in order to characterize the maximum variability that can be expected from different insertions. This is however not likely to be a real-life scenario, as most users will tend to have a position that they find most comfortable and will always try to insert the AHS in their ears in that same position (more or less). Thus, in real-life and for a given user, we expect the variability to be lower than the one measured in this experiment.

This allows us to conclude that although it might seem that positioning of the AHS in the ear could have a strong effect on the sound level at the eardrum (stronger than with a typical supra-aural headset, where positioning is most of the time exactly the same), we can see that the AHS performs in a range of relatively precise and reproducible values.

Intra-headset variability

As the AHS is a consumer product, and not subject to the same quality control requirements as medical devices, it is important to characterize what variability can be expected from different instances of the same model of AHS. In order to do this, we took 4 instances of the AHS, and measured the output level of 3 insertions of each of them on the same setup as in the previous experiment. This time, though, the insertions have been made in a way that tried to minimize the variability due to the position in the artificial ear. Results can be seen in Figure 3 and Figure 4 hereafter.

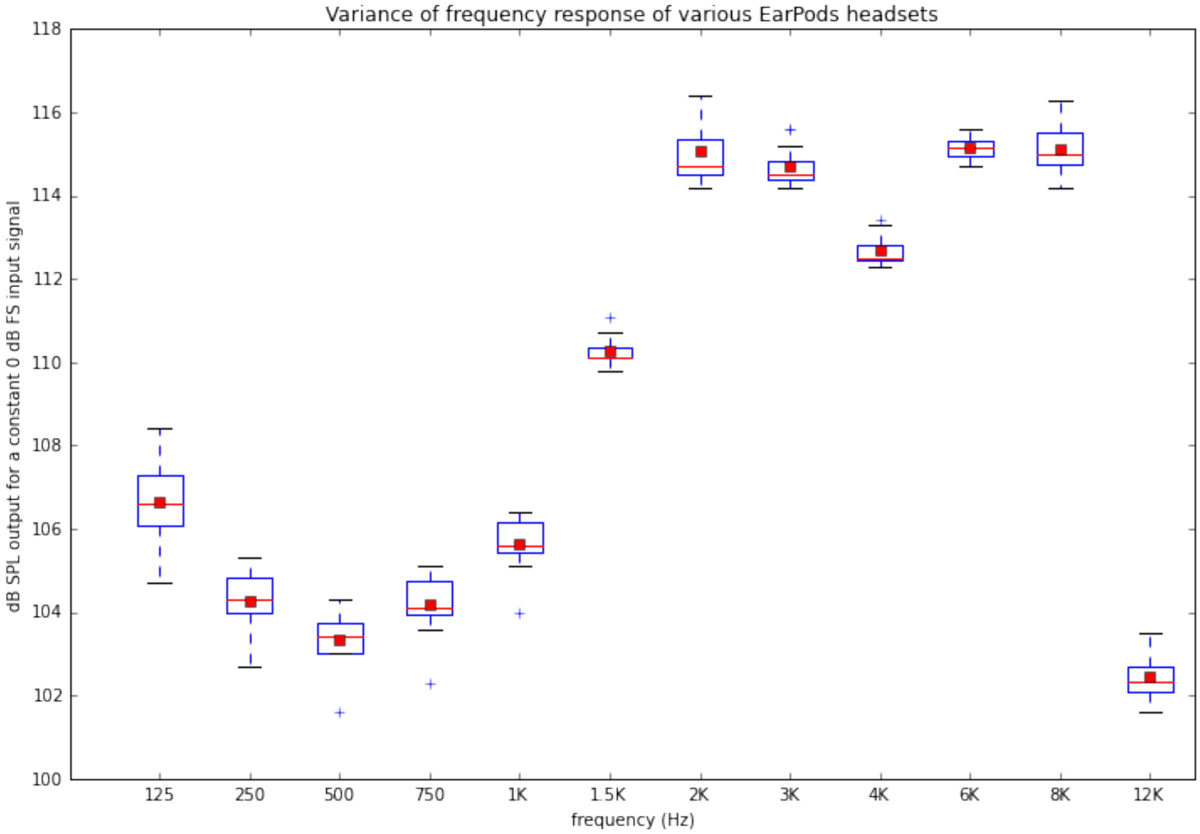

Figure 3: SPL output of various AHS instances in function of frequency

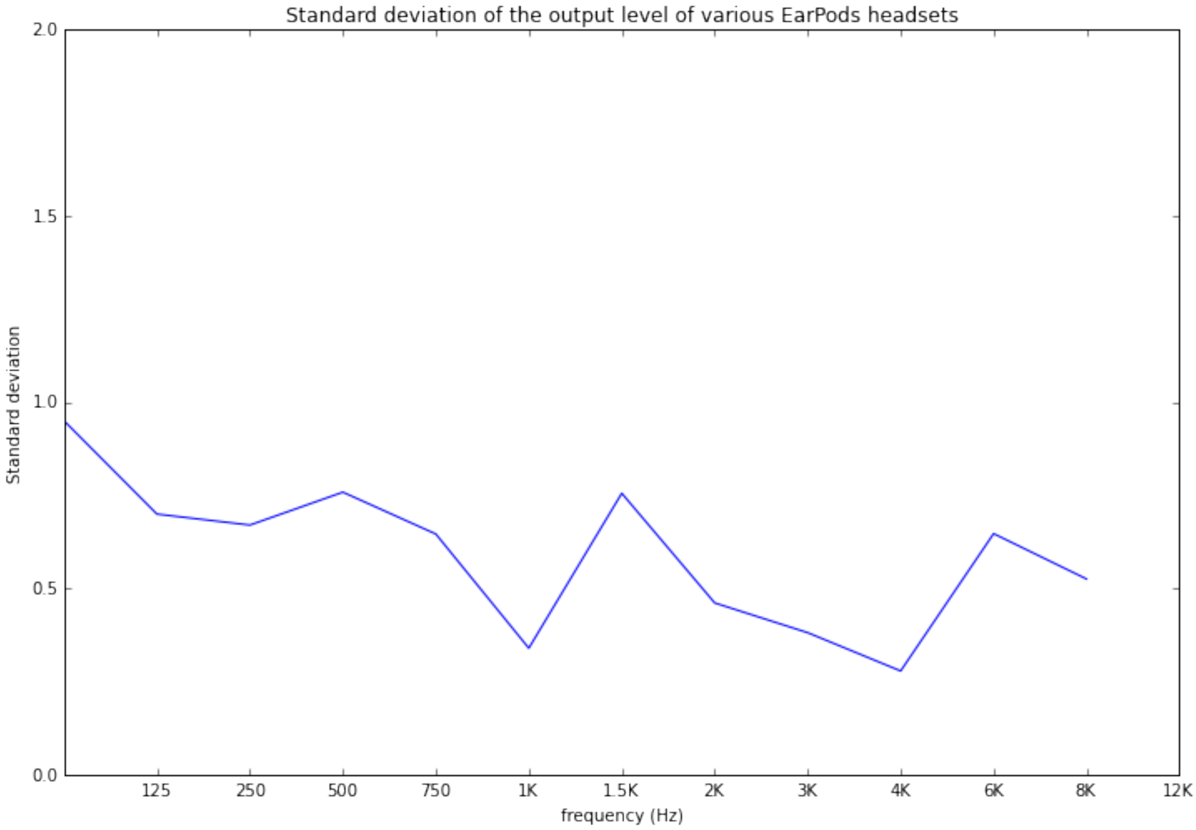

Figure 4: Standard deviation of the SPL output of various AHS instances

As can be seen in the previous plots, especially in fig.4, the variance due to different AHS instances is very small, especially if we take into account the fact those AHS had to be inserted manually in the artificial ear, and although care had to be taken to insert them in the same position, part of this variance can be explained by the difference of fit (and not only by the difference in the transducers' response). In this light, we can expect the standard deviation of the response of the AHS to be on average 0.5dB or less over all the frequency spectrum.

Bonus: Intra-device variability

The DuoTone algorithm is going to run on a system which will consist in: Apple device + AHS + software running the DuoTone algorithm. Thus, we should consider to get calibration values for each supported Apple device + AHS to be able to characterize any supported configuration. In this experiment, we have thus run the same tests as in the previous one, using the same setup for measurement. This time, the same AHS has been left in the test fixture, and we have measured the output response to our test signals using different Apple devices: an iPhone 5, an iPhone 5S, and 2 iPod Touch 5th generation.
The results can be seen in figure 5 and 6 hereafter.

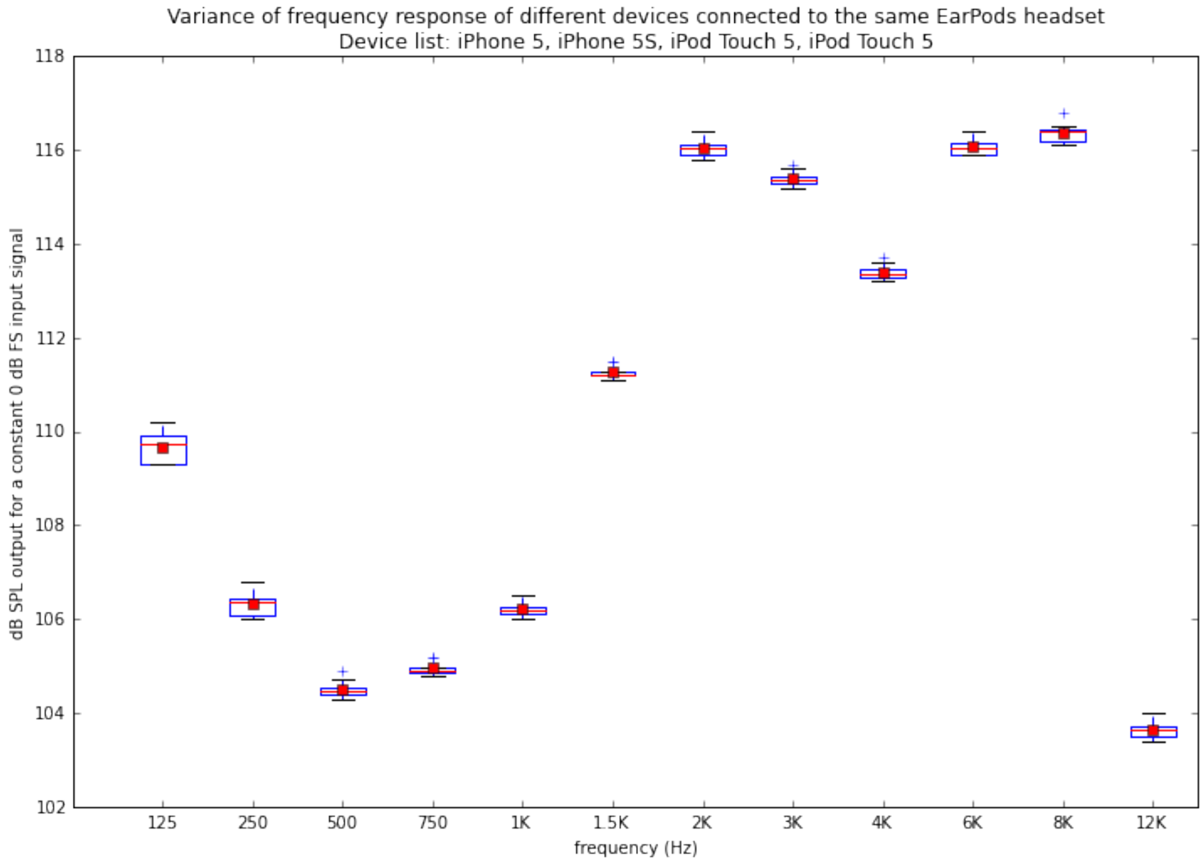

Figure 5: SPL output of various Apple devices connected to the same AHS in function of frequency

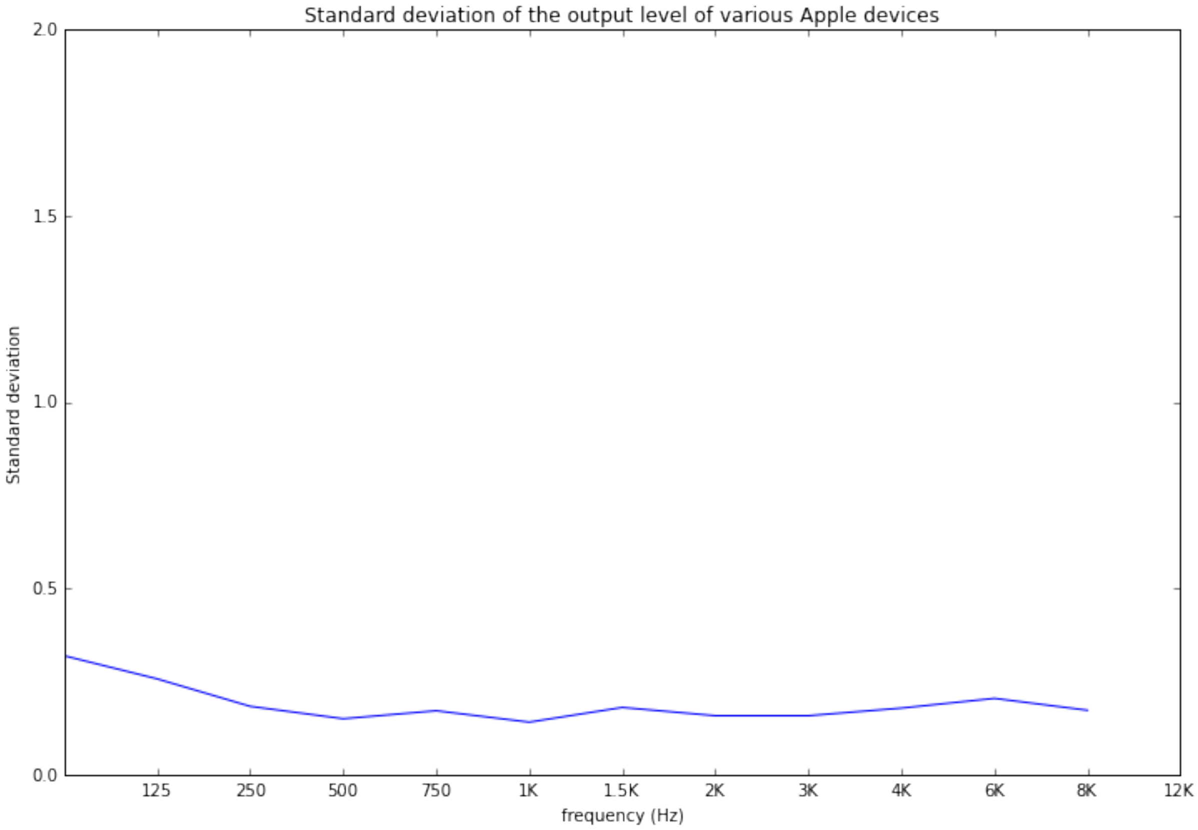

Figure 6: Standard deviation of the SPL output of various Apple devices

We can see from the previous plots that the choice of Apple device has even less influence on the resulting output than the choice of AHS, with a standard deviation of around 0.2 dB. This indicates that we can use any Apple device (from the above mentioned list) indifferently and expect to have the same output. This data will of course need to be updated when new Apple devices come out, in order to ensure that it stays the same, but with the current list of devices we can safely claim that the SPL output is only dependent on the headset, and not the device in use.