PERCEPTION & MEASUREMENT OF HEADPHONE SOUND QUALITY: IS THERE A PREFERRED TARGET RESPONSE?

SEAN OLIVE, TODD WELTI, OMID KHONSARIPOUR

OCTOBER 16 2017
BREEBART (2017) STUDY ON HEADPHONE FREQUENCY RESPONSE VS. RETAIL PRICE

- Measured frequency response of 283 headphones including AE, OE and IE types
- No correlation between frequency response and price
- Evidence that the Headphone Industry has no idea on what the ideal frequency response should be, how to measure it, or how to achieve it?
• The average frequency response of 283 headphones compared to preferred Harman target response for AE headphones

• The Root Mean Square Error varies from 2.5 to 13 dB

• No correlation between price and frequency response
OUR MAIN RESEARCH GOALS

• What is the optimal headphone response in terms of sound quality?
• Is it the same for Around Ear (AE) / On Ear (OE) and In-ear (IE) headphones?
• To what extent do listeners agree on the preferred target response?
• Can we predict listeners’ headphone sound quality preferences based on objective measurements?
Improved Measurement of Leakage Effects for Circum-aural and Supra-aural Headphones

Improved sound quality is achieved when the sound delivered by earphones is quantitatively identical to the sound reproduced by an ear-level loudspeaker in a reference listening room. Listeners rated the same as the actual headphones: the agreement in terms of Pearson correlation *r* = 0.98.

Factors that Influence Listeners’ Preferred Bass and Treble Balance in Headphones

In this study, we present new evidence that the perceived low frequency response. Reducing this lead to a profound effect on preferred bass and treble balance for younger and more experienced listeners.

The Preferred Low Frequency Response of In-Ear Headphones

The preferred target response for In-Ear Headphones was repeated for both loudspeaker and headphone calibrations. Listeners preferred bass and treble level and frequency of the actual and virtualized versions of the IE headphones.

Validation of a Virtual In-ear Headphone Listening Test Method

A new method to evaluate the perceived low frequency response of real In-Ear Headphones was developed by the Rebalanced technique. Listeners chose a different evaluation method for all headphones. Listeners were asked to rate the perceived low frequency response of the actual and virtualized versions of the IE headphones.
The Influence of Program Material on Sound Quality Ratings of In-Ear Headphones

Sean E. Olive, Todd Welti, and Omid Khonsaripour
Harman International, 8500 Balboa Blvd., Northridge, CA, 91329, USA

ABSTRACT
A listening test was conducted to identify music programs that provide sensitive, discriminative, and reliable sound quality ratings for in-ear (IE) headphones. Ten trained listeners gave sound quality ratings for eight models of IE headphones using ten different music programs. A virtualized headphone method was used to provide double-blind, controlled presentations in which headphone leakage effects were monitored and eliminated. The main effect on the sound quality ratings was due to headphones while the program produced no significant effects or interactions. However, certain programs produced more discriminating and reliable ratings than other programs, the key factor being the bandwidth of the program’s spectral content, and the subject’s familiarity with it. As expected, the amount of bass content in each program tended to influence the ratings of headphones that had too much or too little bass output in their measured frequency response.

• 14 paper publications
• Two book chapters on headphone research
• 4 patents pending
• 1-click headphone measurement that predicts sound quality based on machine learning
PRACTICAL BENEFITS OF A PREDICTIVE MODEL

• Eliminates need to conduct time-consuming and expensive listening tests to validate headphone design & performance

• Objective metrics enables consistent sound signature for brand

Applications include:

• headphone design validation/measurement

• Manufacturing Q&A
These are some of the headphones we’ve tested in the last 2-3 years… 100+ models. Do you recognize any of them?
We’ve also tested over 60 in-ear models
## HEADPHONE TYPES

<table>
<thead>
<tr>
<th>Around -Ear</th>
<th>On-Ear</th>
<th>Ear Bud</th>
<th>In-Ear</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Closed-back" /></td>
<td><img src="image2.png" alt="Open-back" /></td>
<td><img src="image3.png" alt="Ear Bud" /></td>
<td><img src="image4.png" alt="In-Ear" /></td>
</tr>
</tbody>
</table>

- **Closed-back**
- **Open-back**

- **Wireless**
- **ANC**
The Relationship between Perception and Measurement of Headphone Sound Quality

Sean E. Olive\textsuperscript{1} and Todd Welti\textsuperscript{2}

\textsuperscript{1} Harman International, Northridge, CA, 91329, USA
sean.olive@harman.com

\textsuperscript{2} Harman International, Northridge, CA, 91329, USA
todd.welti@harman.com

ABSTRACT
Double-blind listening tests were performed on six popular circumaural headphones to study the relationship between their perceived sound quality and their acoustical performance. In terms of overall sound quality, the most preferred headphones were perceived to have the most neutral spectral balance with the lowest coloration. When measured on an acoustic coupler, the most preferred phones produced the smoothest and flattest amplitude response, a response that deviates from the current IEC recommended diffuse-field calibration. The results provide further evidence that the IEC 60268-7 headphone calibration is not optimal for achieving the best sound quality.

Do trained listeners agree on what makes a headphone sound good?
## Headphones Tested

<table>
<thead>
<tr>
<th>Brand / Model</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKG K701</td>
<td>$278</td>
</tr>
<tr>
<td>AKG K550</td>
<td>$245</td>
</tr>
<tr>
<td>Audeze LCD2 (rev 2)</td>
<td>$995</td>
</tr>
<tr>
<td>Beats by Dre Studio Limited Edition</td>
<td>$270</td>
</tr>
<tr>
<td>Bose Quiet Comfort 15</td>
<td>$299</td>
</tr>
<tr>
<td>V-Moda Crossfade LP</td>
<td>$115</td>
</tr>
</tbody>
</table>
The influence of visual & psychological biases (e.g. brand, price, appearance and celebrity endorsement) were removed from listeners’ judgement of sound quality.
EVEN THE MOST POPULAR HEADPHONES ARE QUITE DIFFERENT IN TERMS OF THEIR MEASURED AND PERCEIVED SPECTRAL BALANCE

measured response in GRAS ear simulator
perceived spectral balance by trained listeners
Improved Measurement of Leakage Effects for Circum-aural and Supra-aural Headphones

Todd Wettl
Harman International Inc. Northridge CA, 91360, USA
todd.wettl@harman.com

ABSTRACT

Headphone leakage effects can have a profound effect on low frequency performance of headphones. A large survey, including over 2000 individual headphone measurements, was undertaken in order to compare leakage effects on test subjects and leakage effects of the same headphones measured on a test fixture. Ten different commercially available headphones were used, each measured on eight different test subjects and a test fixture with several sets of pinnae. Modifications to the pinnae were investigated to see if the leakage effects measured on the test fixture could be made to better match the real world leakage effects measured on human test subjects.
LEAKAGE PROBLEMS WITH STANDARD IEC PINNA

KB0070 IEC 60318-4 standard pinna on G.R.A.S 45CA

large gap

headphone leakage
Measure 10 headphones on 8 subjects

Repeat measurements on Flat Plate & GRAS 45 CA different sets of standard and modified pinnae

WHICH MEASUREMENT BEST MATCHES HUMAN MEASUREMENTS ??
Standard IEC pinnae was modified to reduce protrusion, reduce gap between Helix and Tragus and reduce stiffness.
Plate under estimates leakage

IEC Mod2 produces least error

IEC Pinna over estimates leakage

Figure 13  Same data as Figure 12, but averaged over all headphones. Average absolute value of error from 50-500 is shown.
Since this paper was written, G.R.A.S has released a new pinna for KEMAR & G.R.A.S 45 CA that more accurately measures leakage effects of humans.

- It has no protrusion and has a softer/less stiff compliance similar to the human ear.

**Question:**

- *Do we need to measure AE headphones on artificial heads to accurately produce leakage? Which artificial head is most accurate?*
Is there a more preferred headphone target curve response than the standard DF (diffuse field) and FF (free-field) calibrations?

Listener Preference For Different Headphone Target Response Curves

Sean E. Olive¹, Todd Welt², and Elisabeth McMullin³

Harman International, Northridge, CA, 91329, USA

¹ sean.olive@harman.com
² todd.welt@harman.com
³ elisabeth.mcmullin@harman.com

ABSTRACT

There is little consensus among headphone manufacturers on the preferred headphone target frequency response required to produce optimal sound quality for reproduction of stereo recordings. To explore this topic further, we conducted two double blind listening tests in which trained listeners rated their preferences for 8 different headphone target frequency responses reproduced using two different models of headphones. The target curves included the diffuse-field and free-field curves in ISO 1996-2, a modified diffuse-field target recommended by LeCha, the unequalized headphone, and a new target response based on acoustic measurements of a calibrated loudspeaker system in a listening room. For both headphones, the new target based on the in-room loudspeaker response was the most preferred headphone target response curve.
Our simple logic is as follows:

Since stereo recordings are optimized to sound best through loudspeakers in rooms...

.. stereo recordings will sound best when reproduced through headphones that simulate the in-room response of a well-designed loudspeaker system calibrated in a reference listening room.
HARMAN REFERENCE LISTENING ROOM (2013)

Measuring the in-room loudspeaker response
# HEADPHONE TARGET CURVES TESTED

<table>
<thead>
<tr>
<th>Equalization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>No EQ</td>
<td>Headphone with no EQ</td>
</tr>
<tr>
<td>DF_MH</td>
<td>Diffuse-field target based on Hammershöi &amp; Möller [7]</td>
</tr>
<tr>
<td>DF_M</td>
<td>Diffuse-field target response based on Möller [8]</td>
</tr>
<tr>
<td>DL_L</td>
<td>A modified diffuse-field calibration based on Lorho [4]</td>
</tr>
<tr>
<td>FF</td>
<td>A free-field calibration based on Hammershöi &amp; Möller [7]</td>
</tr>
<tr>
<td>Harman Target 1</td>
<td>Based on measurements of JBL Pro LSR loudspeakers in Harman Reference Room</td>
</tr>
<tr>
<td>Harman Target 2</td>
<td>Same as above with modified in-room target curve with slightly less bass and treble</td>
</tr>
</tbody>
</table>
LISTENING TEST DESIGN

Sennheiser HD 518 ($120)

Audeze LCD2 ($995)

Each target curve was rated by trained listeners based on preference using three music programs with one repeat.

The test was repeated using two different headphones equalized to the different target responses.
RESULTS
PREFERRED HEADPHONE TARGET RESPONSE

<table>
<thead>
<tr>
<th>Headphone Target Curve</th>
<th>Preference Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>5.85</td>
</tr>
<tr>
<td>DF_M</td>
<td>3.77</td>
</tr>
<tr>
<td>DF_MH</td>
<td>3.35</td>
</tr>
<tr>
<td>DF_L</td>
<td>3.28</td>
</tr>
<tr>
<td>No EQ</td>
<td>2.88</td>
</tr>
<tr>
<td>FF</td>
<td>1.32</td>
</tr>
</tbody>
</table>
PREFERRED HEADPHONE TARGET RESPONSE

<table>
<thead>
<tr>
<th>Headphone Target Response</th>
<th>Preference Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2</td>
<td>7.03</td>
</tr>
<tr>
<td>T1</td>
<td>4.83</td>
</tr>
<tr>
<td>DF_M</td>
<td>4.75</td>
</tr>
<tr>
<td>No EQ</td>
<td>4.08</td>
</tr>
<tr>
<td>DF_MH</td>
<td>2.27</td>
</tr>
<tr>
<td>FF</td>
<td>1.39</td>
</tr>
</tbody>
</table>

©2016 HARMAN INTERNATIONAL INDUSTRIES, INCORPORATED
The Influence of Listeners’ Experience, Age, and Culture on Headphone Sound Quality Preferences

Sean E. Oliver1, Todd Welti2, and Elisabeth McMullin2
Harman International, Northridge, CA, 91329, USA

1 Sean.Oliver@harman.com
2 Todd.Welti@harman.com
2 Elisabeth.McMullin@harman.com

ABSTRACT

Double-blind headphone listening tests were conducted in four different countries (Canada, USA, China, and Germany) involving 238 listeners of different ages, gender and listening experiences. Listeners gave comparative preference ratings for three popular headphones and a new reference headphone that were virtually presented through a common replicator headphone equalized to match their measured frequency responses. In this way, biases related to headphone brand, price, visual appearance and comfort were removed from listeners’ judgment of sound quality. On average, listeners preferred the reference headphone that was based on the in-room frequency response of an accurate loudspeaker calibrated in a reference listening room. This was generally true regardless of the listener’s experience, age, gender and culture. This new evidence suggests a headphone standard based on this new target response would satisfy the tastes of most listeners.
The Beats With a Billion Eyes

He’s conquered the headphones market, but Dr. Dre isn’t selling great sound. He’s not even selling celebrity. He’s selling the concept of “bass.”

Jesse Dorris. Slate Magazine Sept. 11 2013

Is their success more about the marketing than the sound (bass)?
HEADPHONE VIRTUALIZATION METHOD

1. Target Headphone
   - Measure at DRP
2. Simulator Headphone
   - Measure at DRP
3. Design Virtual HP Filter (IIR)
4. Apply Virtual HP Filter to Simulator Phone
5. Flatten EQ
Compared spectral and preference ratings of 7 actual versus virtual headphones

Good agreement in \( r = 0.86 \) in ratings

Some notable differences that were specific to 1-2 models and individual differences

Likely explained by leakage effects (actual headphones) and biases related to headphone fit/weight.

INTRODUCTION
Comparative listening tests on headphones are challenging to conduct in a controlled, double-blind fashion. With some effort, the highest variable - human evaluation - is virtually impossible to remove from the test. Moreover, blind comparative headphone listening tests require the test administrators to manually substitute the different headphones on the subject's head over several trials, making it an extremely tedious, invasive, and distracting exercise for both the listener and administrator.

A VIRTUAL HEADPHONE LISTENING TEST METHODOLOGY
SEAN E. OLIVE†, TODD WEILITZ, AND ELISABETH MCMULLIN‡

Harman International, Northridge, CA, USA
sean_olive@harman.com
todd.wellt@harman.com
elizabeth.mcmullin@harman.com

Comparative listening tests on multiple headphones are challenging to conduct in a controlled, double-blind fashion. One solution is to present the listener virtualized versions of the headphones through a single reference headphone that is equalized to simulate the three-dimensional response of the different headphones under test. This paper describes a method for conducting virtual headphone listening tests and presents results of a validation experiment where listener sound quality ratings from standard and virtual headphone listening tests are compared. The listening test results show good correlation between the two methods in terms of perceived spectral balance and overall preference.

©2016 HARMAN INTERNATIONAL INDUSTRIES, INCORPORATED
WHAT ABOUT CULTURAL DIFFERENCES IN TASTE IN HEADPHONE SOUND QUALITY?

- Canada (untrained)
- United States (trained vs untrained)
- China (trained vs untrained)
- Germany (trained vs untrained)
## HEADPHONES TESTED

<table>
<thead>
<tr>
<th>Brand / Model</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harman Target Curve</td>
<td>Based on latest AES paper October 2013</td>
</tr>
<tr>
<td>Sennheiser HD800</td>
<td>$1500</td>
</tr>
<tr>
<td>Audeze LCD2 (rev 2)</td>
<td>$995</td>
</tr>
<tr>
<td>Beats by Dre Studio Limited Edition</td>
<td>$270</td>
</tr>
<tr>
<td>Group</td>
<td>Country</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td>Harman NR</td>
<td>USA</td>
</tr>
<tr>
<td>Harman FH</td>
<td>USA</td>
</tr>
<tr>
<td>Harman KB</td>
<td>Germany</td>
</tr>
<tr>
<td>Harman SZ</td>
<td>China</td>
</tr>
<tr>
<td>Citrus College</td>
<td>USA</td>
</tr>
<tr>
<td>LMU</td>
<td>USA</td>
</tr>
<tr>
<td>Harris Insitute</td>
<td>Canada</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
</tr>
<tr>
<td>Program / Artist / CD / Track / Label / Release Date</td>
<td>Details</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>ES</strong> / Estelle w. Kayne West / Shine / American Boy / Atlantic Records, 2008, B00142Q7H8 / April 29 2008</td>
<td>Hip Hop with male and female vocal</td>
</tr>
<tr>
<td><strong>SD</strong> / Steely Dan / Two Against Nature / Cousin Dupree / Giant Records / B00004GOXS / February 29 2000</td>
<td>Jazz Pop with Male Vocal</td>
</tr>
<tr>
<td><strong>JW</strong> / Jennifer Warnes / Blue Raincoat / Bird on a Wire</td>
<td>Rock Pop with Female Vocal</td>
</tr>
</tbody>
</table>
Listeners rate headphones A through D based on preference and give comments (optional).
**LISTENING EXPERIENCE EFFECT ON HEADPHONE PREFERENCE**

**Conclusion:**
Listening experience has little influence on headphone preference.
Conclusion:
Listeners preferred the same headphones regardless of which country they live in.
AGE CATEGORIES OF LISTENERS IN YEARS

- <25: 37%
- 25-34: 30%
- 35-44: 18%
- 45-54: 13%
- >55: 3%

Gender Distribution:
- Male: 50%
- Female: 50%
Presented is a graph illustrating the effect of listener's age on headphone preference. The x-axis represents different age groups: <25, 25-34, 35-44, 45-54, >55. The y-axis represents the preference rating ranging from 1 to 7.

The graph shows the preference ratings for four different headsets labeled HP1, HP2, HP3, and HP4. The ratings are depicted as lines with error bars indicating variability.

**Conclusion:**
Listeners’ age has little influence on headphone preference.
MEAN PREFERENCE RATINGS (N=238 LISTENERS)

Harman Target Curve

<table>
<thead>
<tr>
<th>Preference Rating</th>
<th>HP2</th>
<th>HP3</th>
<th>HP4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.8</td>
<td>5.0</td>
<td>2.2</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

virtual HP = dotted curve
actual HP = solid curve
Factors that Influence Listeners’ Preferred Bass and Treble Balance in Headphones

Sean E. Olive\(^1\) and Todd Welti\(^2\)
Harman International, Northridge, CA, 91329, USA
\(^1\)sean.olive@harman.com \(^2\)todd.welti@harman.com

ABSTRACT

A listening experiment was conducted to study factors that influence listeners’ preferred bass and treble balance in headphone sound reproduction. Using a method of adjustment, a total of 249 listeners adjusted the relative treble and bass levels of a headphone that was first equalized at the eardrum reference point (DRP) to match the in-room steady-state response of a reference loudspeaker in a reference listening room. Listeners repeated the adjustment five times using three stereo music programs. The listeners included males and females from different age groups, listening experiences, and nationalities. The results provide evidence that the preferred bass and treble balances in headphones were influenced by several factors including program, and the listeners’ age, gender and prior listening experience. The younger and less experienced listeners on average preferred more bass and treble in their headphones compared to the older, more experienced listeners. Female listeners on average preferred less bass and treble than their male counterparts.

1. INTRODUCTION

Recent scientific investigations into alternative headphone target curves have found that listeners prefer them when compared to the standard diffuse and free-field headphone calibrations \(1\)–\(4\). Olive et al. showed evidence that trained listeners preferred a headphone target response that closely matched the measured in-room stable-state response of a reference loudspeaker \(3\). However, the relative levels of the bass and treble sections of the headphone target response were derived empirically rather than through formal experimentation, leaving some doubt as to whether the bass and treble levels of the headphone target response were optimized for best sound quality.

To address this question, a follow up experiment was recently conducted wherein listeners directly adjusted the relative bass and treble levels of the headphone after it was equalized at the DRP to match the in-room stable-state response of a reference loudspeaker. The experiment was repeated for both loudspeaker and headphone playback conditions to determine how closely the two results matched. The average preferred bass and treble levels were 4.8 dB and -4.4 dB, respectively, for headphone playback, and 6.6 dB and -2.4 dB for loudspeaker reproduction. In other words, listeners adjusted the relative bass and treble levels of the headphone to match the preferred balance of the loudspeaker. The results provide evidence that the preferred bass and treble balances in headphones were influenced by several factors including program, and the listeners’ age, gender and prior listening experience. The younger and less experienced listeners on average preferred more bass and treble in their headphones compared to the older, more experienced listeners. Female listeners on average preferred less bass and treble than their male counterparts.

Do the different listeners like the same amount of bass and treble levels in their headphones?
LISTENING TEST SETUP

Task: Using the two rotating knobs on the iPad adjust the bass and treble levels according to personal preference

Sennheiser HD518 equalized to “flat”
Harman Target Response

Custom iPad Test Software
HARMAN HEADPHONE IPAD TEST SOFTWARE

Bass Control

Treble Control

©2016 HARMAN INTERNATIONAL INDUSTRIES, INCORPORATED
LISTENING TEST METHOD

Using a Method of Adjustment, listeners adjust the bass and treble levels of the headphone in 0.25 dB increments according to personal preference.

Repeat 5x using three programs.

Randomize starting level(s) and program order.
<table>
<thead>
<tr>
<th>Group</th>
<th>Country</th>
<th>Sample Size (n)</th>
<th>Gender</th>
<th>Age of Listeners in Years</th>
<th>Listening Experience % of Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Country</td>
<td>Sample Size (n)</td>
<td>Gender</td>
<td>Median</td>
<td>SD</td>
</tr>
<tr>
<td>Harman NR</td>
<td>USA</td>
<td>9</td>
<td>78%</td>
<td>39.4</td>
<td>10.4</td>
</tr>
<tr>
<td>Harman FH</td>
<td>USA</td>
<td>24</td>
<td>83%</td>
<td>36.0</td>
<td>12.9</td>
</tr>
<tr>
<td>Harman KB</td>
<td>Germany</td>
<td>72</td>
<td>93%</td>
<td>38.0</td>
<td>9.1</td>
</tr>
<tr>
<td>Harman SZ</td>
<td>China</td>
<td>20</td>
<td>80%</td>
<td>31.5</td>
<td>6.0</td>
</tr>
<tr>
<td>Citrus College</td>
<td>USA</td>
<td>24</td>
<td>76%</td>
<td>23.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Loyola M. University (LMU)</td>
<td>USA</td>
<td>15</td>
<td>93%</td>
<td>21.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Harris College</td>
<td>Canada</td>
<td>48</td>
<td>90%</td>
<td>23.0</td>
<td>6.0</td>
</tr>
<tr>
<td>JBL Dealers</td>
<td>Varied</td>
<td>37</td>
<td>100%</td>
<td>40.0</td>
<td>9.3</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>249</td>
<td>89%</td>
<td>32</td>
<td>11.4</td>
</tr>
</tbody>
</table>
RESULTS
Preferred Bass and Treble Levels

<table>
<thead>
<tr>
<th>Relative Level (dB)</th>
<th>Bass</th>
<th>Treble</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bass and Treble Filters Gain

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Relative Level (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>6.44 dB</td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td></td>
</tr>
</tbody>
</table>

-1.41 dB
### Conclusion: The amount of preferred bass and treble is program-dependent likely due to “circle-of-confusion” issues.
EFFECT OF GENDER ON PREFERRED BASS AND TREBLE LEVELS

Conclusion: Females like less bass and treble than males. More research is needed to confirm this.
Conclusion: Listeners with more experience in critical listening prefer less bass and treble than listeners with less experience.
EFFECT OF AGE OF LISTENER

Conclusion: With increasing age, listeners prefer less bass and less treble until 56+ years where they prefer much less bass and more treble possibly due to hearing loss.
Conclusion: Similar preferences with exception of German listening panel. More research needed.
LISTENING GROUPS

Bass

Treble

Moderate Bass & Treble Lovers

Heavy Bass & Treble Lovers
COMPARISON WITH PREVIOUS STUDIES

Playback: Headphone                   Headphone                         Loudspeaker

- Olive & Welti (2015) 6.6
- Olive et al (2013) 4.8
- Olive et al (2014) 6.6

Olive & Welti (2015)
Olive et al (2013)
Olive et al (2014)
Preferred headphone target curve, the DF calibration would make the headphone sound too thin and bright due to the lower bass and higher treble levels. This was reported in a previous study [6], and has been confirmed again in the current study.

Figure 17 The preferred headphone target response measured at DRP (black) based on this study. Also shown is the measured response of the loudspeaker equalized to a flat in-room target response.

4.3 Measured Responses of Loudspeaker Equalized to the Preferred Target

Fig. 18 shows the measured frequency response of the Revel F208 loudspeaker equalized to the preferred in-room target response (solid curve) found in this study.

For the sake of comparison, we also show the measured response of the loudspeaker equalized to a flat in-room target response. The flat in-room loudspeaker response curve would have too much treble and not enough bass to produce satisfying results for listener as confirmed in this study and a previous one [5].

Finally, Fig. 19 shows the measured in-room response of the Revel F208 (solid line) equalized to the preferred target response curve. This measured response is very similar to the in-room loudspeaker target (dotted) that was preferred by listeners in two previous studies where they evaluated different loudspeaker-room correction products [5], [6]. Also shown in Fig. 19 is the predicted in-room response of the loudspeaker based on anechoic measurements (see Fig. 1). Above 200 Hz there is good agreement between the predicted in-room response of the loudspeaker (based on anechoic measurements with no room equalization), and the measured in-room response of the loudspeaker equalized to the preferred target response.

What this tells us is that a well-designed loudspeaker shouldn’t require much equalization above the transition frequency where the room no longer dominates the quality of sounds heard. However, below 100–300 Hz, the loudspeaker will likely need equalization to deal with room mode and boundary effects, and possibly some bass enhancement to satisfy the tastes of individual listeners, and accommodate variations in the quality of program material.

Figure 19 The measured in-room response of the Revel F208 loudspeaker equalized to the preferred in-room target curve (black), the predicted in-room response of the loudspeaker (cyan) based on anechoic measurements (see Fig. 1), and the modified in-room loudspeaker target curve, RR1 (red dotted) from [6].

Olive et al. Headphone Target Curve from [4]

Loudspeaker equalized to flat in-room response [4]

SOME LIMITATIONS OF RESEARCH

- Method of Adjustments in bass and treble not loudness compensated: some listeners may have simply boosted bass and treble because it made the music louder
- Listener sample was not randomly selected or balanced
- Listener training was self-reported (exception was Harman trained listeners)
- Programs limited to 3 samples
- Audiometric hearing performance was not tested (except Harman trained listeners)
- In practice, preferred bass and treble level may vary with background noise (masking) and headphone fit (leakage)
The *ideal* headphone target curve is a **moving target** due to variations in:

- Program material ("circle of confusion" / lack of meaningful standards in recording industry monitoring chain)
- Individual taste (age and experience seem to matter)
- Masking effects of background noise (bass/mids are masked)
- Variation in headphone fit (bass leakage)

We need a simple bass / treble control on headphones to compensate for these variations
The Preferred Low Frequency Response of In-Ear Headphones

Sean E. Olive, Todd Welti, and Omid Khonsaripour
Harman International, 8500 Balboa Blvd., Northridge, CA, USA, 91329
Correspondence should be addressed to Sean Olive (sean.olive@harman.com)

ABSTRACT
A series of controlled listening tests were conducted to determine the preferred low frequency response of in-ear (IE) headphones. Using a method of adjustment ten trained listeners adjusted the bass level and frequency of a 2nd order low shelving filter applied to a high quality IE headphone that was calibrated to the preferred target response of a circumaural headphone [5]. The adjustments were done for three different music programs, and repeated with and without loudness normalization and control of leakage effects. The influence of program, individual taste, and loudness normalization and leakage effects on preferred low frequency response are presented and discussed.

1 Introduction

were made without loudness normalization, and whether this influenced the results is something that is addressed in the current study.

There are few reported studies on the preferred target response of IE headphones even though these types of headphones represent the largest segment of headphones sold [1]. An important research question is whether the IE target response should be the same as the OE target response, and if not, why? Possible reasons could be related to effects of the occluded ear, low frequency leakage effects on bass performance, and the absence of pinnae effects that are present in OE headphones but not IE types.

This paper reports some listening experiment designed to answer the following research questions:

• What influence does loudness have on the preferred low frequency response of IE headphones?
• What influence does ear leakage have on the preferred low frequency response of IE headphones?
• What influence does ear leakage have on the preferred low frequency response of IE headphones, with and without loudness normalization?

Do listeners prefer the same target response for IE headphones as around-ear (AR) and on-ear headphones?
• We don’t currently know what the preferred target response for IE headphones should be.....

• The preferred target responses for AE and OE headphones may not be applicable for IE headphones because:
  
  • No contribution from out-ear/pinnae effects
  • Occluded ear produces unique acoustic effects
  • Low frequency response (< 500 Hz) of IE headphones significantly impacted by fit/seal/ leakage effects
LISTENING EXPERIMENTS

- 10 trained listeners adjusted the Level (dB) and Frequency (Hz) of a 2nd order LF shelf filter
- Task repeated 6 times using 3 music programs (18 trials)
- Starting levels and frequencies randomized
- Four tests where the influence of loudness and leakage were studied
RANGE OF LEVEL & FREQUENCY SETTINGS

- Level adjustments in 0.25 dB increments over a 30 dB range (-5 to +25 dB)
- Frequency adjustment were available between 50 Hz to 1kHz in logarithmic increments
## Listening Experiments

<table>
<thead>
<tr>
<th>Test</th>
<th>One</th>
<th>Two</th>
<th>Three</th>
<th>Four</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Loudness Normalization</strong></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Leakage Control</strong></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Program / Track / Album / Label / Release Date</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES - Estelle w. Kayne West / Shine / American Boy / Atlantic Records, 2008, B00142Q7H8 / April 29 2008</td>
<td>Hop Hop with male and female vocal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD - Steely Dan / Two Against Nature / Cousin Dupree / Giant Records / B00004GOXS / February 29 2000</td>
<td>Jazz Pop with Male vocal</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
LISTENING TEST SETUP

Mac with custom MAX/MSP Patch

GRiffin USB PowERMate (Bass level)

GRiffin USB PowERMate (Freq)

Passive Switcher

Leakage Measurement
SUBJECT 450: FIVE ATTEMPTS BEFORE GOOD SEAL
RESULTS
**Conclusion:** Test One produced slightly higher preferred bass level settings than the other three tests.
Conclusion: Test Four produced lower preferred frequency settings than Tests One and Three
Based on average of all four tests.
TARGET CURVES FOR IE VERSUS AE HEADPHONES

Current Study IE Headphones
Olive et al. (2013) AE Headphones
Olive et al. (2015) AE Headphones
• Most commercial IE headphones have target responses that deviate significantly from what our trained listeners prefer

• There is an opportunity to significantly improve the sound quality of IE headphones
Validation of a Virtual In-ear Headphone Listening Test Method

Todd Welti\textsuperscript{1}, Sean E. Olive\textsuperscript{2}, and Omid Khonsaripour\textsuperscript{3}

Harman International, Northridge, California, USA

Email: \textsuperscript{1}todd.welti@harman.com, \textsuperscript{2}sean.olive@harman.com, \textsuperscript{3}omid.khonsari@harman.com

Correspondence should be addressed to Todd Welti (todd.welti@harman.com)

ABSTRACT

Controlled, comparative double blind listening tests on different in-ear (IE) headphones are logistically challenging to conduct. One solution is to present listeners with virtualized versions of the headphones through a high quality IE replicator headphone equalized to match their measured frequency responses. To test the accuracy of the virtual headphone method, ten trained listeners evaluated the overall sound quality of both the actual and virtualized versions of twelve different IE headphones that were binaurally recorded on a standard coupler and reproduced through a calibrated replicator headphone. The results show the different models of headphones produced the main effect on perceived sound quality. The virtualized headphones were essentially rated the same as the actual headphones: the agreement in terms of Pearson correlation was \( r = 0.98 \).

1 Introduction

Scientific listening tests on headphones are difficult to conduct owing to the challenges in controlling listening test nuisance variables and their inherent biases. They include sighted and tactile biases, among the different headphones in a controlled and double blind fashion. Hiroven at al. \cite{Hiroven2016} employed this aurialization method to evaluate six different headphones (four circumaural and two intra-concha models) using narrow and wide-band speech. The results were notable differences in agreement between the different test methods, \( \text{correlation} = 0.98 \), which they attributed to sighted biases, and errors related to leakage and variations in fit during the recording and reproduction stages.

But first, we had to determine if it produced accurate and valid results.
## IE HEADPHONES TESTED

<table>
<thead>
<tr>
<th>TEST ONE</th>
<th>TEST TWO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Anchor (AKG Y20)</td>
<td>Anchor (AKG Y20)</td>
</tr>
<tr>
<td>Shure Se215</td>
<td>Panasonic RP TCM</td>
</tr>
<tr>
<td>Sony MDR</td>
<td>Polk AM 5110</td>
</tr>
<tr>
<td>AKG K3003</td>
<td>AKG 323 XS</td>
</tr>
<tr>
<td>Klipsch S4</td>
<td>RHA</td>
</tr>
<tr>
<td>Panasonic RP HJE</td>
<td>AKG Y23</td>
</tr>
</tbody>
</table>
Trained listeners gave double-blind ratings for 12 different models of IE headphones using recordings of actual & virtual versions of the headphones.

Hidden reference and anchor included.

Leakage eliminated in tests by monitoring leakage via miniature mic.
MEASUREMENTS OF ACTUAL & VIRTUAL HEADPHONES

Actual Headphone

Virtual Headphone

Highest Rated (below Target)  Lowest Rated
VALIDATION TEST ONE

Correlation = 0.99

Sound Quality Rating

<table>
<thead>
<tr>
<th>Headphone</th>
<th>Actual Headphone</th>
<th>Virtual Headphone</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50 ± 5</td>
<td>50 ± 5</td>
</tr>
<tr>
<td>Reference</td>
<td>50 ± 5</td>
<td>50 ± 5</td>
</tr>
<tr>
<td>C</td>
<td>25 ± 5</td>
<td>25 ± 5</td>
</tr>
<tr>
<td>D</td>
<td>25 ± 5</td>
<td>25 ± 5</td>
</tr>
<tr>
<td>E</td>
<td>25 ± 5</td>
<td>25 ± 5</td>
</tr>
<tr>
<td>F</td>
<td>25 ± 5</td>
<td>25 ± 5</td>
</tr>
<tr>
<td>Anchor</td>
<td>10 ± 5</td>
<td>10 ± 5</td>
</tr>
</tbody>
</table>
VALIDATION TEST TWO

Sound Quality Rating

Actual Headphone
Virtual Headphone

Correlation = 0.95

Headphone
Reference  H  I  J  K  L  Anchor

©2016 HARMAN INTERNATIONAL INDUSTRIES, INCORPORATED
Is there a preferred target curve for IE headphones?

Can we model and predict listener preference ratings based on its acoustic performance?
• Listening tests take too much time and money to conduct and there is pressure to eliminate them or do faster, cheaper, less controlled tests

• Machine learning & AI will eventually replace human listeners for product testing / room tuning / personalization

• To get there we will need lots of good subjective and objective measurement data to train and validate the machines
• A subset of 30 different models from 19 different manufacturers were selected from a sample of 62 models

• Varied in retail price from $26 - $1000.

• Included dynamic and balanced armature designs including single and multi-way designs
<table>
<thead>
<tr>
<th>Headphone Brand / Model</th>
<th>Driver Type ( D = \text{Dynamic} ) ( BA = \text{Balanced Armature} )</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AKG H300</td>
<td>D</td>
<td>N/A</td>
</tr>
<tr>
<td>2. AKG K3003 (Reference Filter)</td>
<td>3-way: ( 1 \text{ D} + 2 \text{ BA} )</td>
<td>$1,000</td>
</tr>
<tr>
<td>3. AKG N20</td>
<td>D</td>
<td>$130</td>
</tr>
<tr>
<td>4. AKG Y20</td>
<td>D</td>
<td>$30</td>
</tr>
<tr>
<td>5. Audio Technica ATH-CKR10</td>
<td>D</td>
<td>$202</td>
</tr>
<tr>
<td>6. B&amp;O BeoPlay H3</td>
<td>D</td>
<td>$150</td>
</tr>
<tr>
<td>7. B&amp;W C5 S2</td>
<td>D</td>
<td>$180</td>
</tr>
<tr>
<td>8. Beats urBeats</td>
<td>D</td>
<td>$63</td>
</tr>
<tr>
<td>9. Beyerdynamic iDX 16iE</td>
<td>D</td>
<td>$134</td>
</tr>
<tr>
<td>10. Bose Quiet Comfort 20</td>
<td>D</td>
<td>$300</td>
</tr>
<tr>
<td>11. Bose SoundSport</td>
<td>D</td>
<td>$100</td>
</tr>
<tr>
<td>12. Etymotic ER4P</td>
<td>BA</td>
<td>$200</td>
</tr>
<tr>
<td>13. Flare</td>
<td>D</td>
<td>$250</td>
</tr>
<tr>
<td>14. Focal Sphear</td>
<td>D</td>
<td>$149</td>
</tr>
<tr>
<td>15. HIFIMAN RE-400</td>
<td>D</td>
<td>$79</td>
</tr>
<tr>
<td>16. JBL Synchros S100</td>
<td>D</td>
<td>N/A</td>
</tr>
<tr>
<td>17. Klipsch R6</td>
<td>D</td>
<td>$79</td>
</tr>
<tr>
<td>18. LG/HK HBS-1100 BT</td>
<td>D</td>
<td>N/A</td>
</tr>
<tr>
<td>19. Philips TX2</td>
<td>D</td>
<td>$40</td>
</tr>
<tr>
<td>20. PSB M4U-4</td>
<td>2-way: ( D + BA )</td>
<td>$300</td>
</tr>
<tr>
<td>21. RHA MA750i</td>
<td>D</td>
<td>$130</td>
</tr>
<tr>
<td>22. Sennheiser IE80</td>
<td>D</td>
<td>$308</td>
</tr>
<tr>
<td>23. Sennheiser Momentum M2IEi</td>
<td>D</td>
<td>$100</td>
</tr>
<tr>
<td>24. Shure SE215</td>
<td>D</td>
<td>$100</td>
</tr>
<tr>
<td>25. Shure SE535</td>
<td>Triple Micro drivers</td>
<td>$500</td>
</tr>
<tr>
<td>26. Shure SE846 (Neutral Setting)</td>
<td>Quad Micro drivers</td>
<td>$990</td>
</tr>
<tr>
<td>27. Sony MDR-7550</td>
<td>D</td>
<td>$230</td>
</tr>
<tr>
<td>28. Sony XBA-H1</td>
<td>2-way: ( D + BA )</td>
<td>$97</td>
</tr>
<tr>
<td>29. Soul by Ludacris</td>
<td>D</td>
<td>$90</td>
</tr>
<tr>
<td>30. Symphonized NRG 2.0</td>
<td>D</td>
<td>$26</td>
</tr>
<tr>
<td>Program</td>
<td>Artist/Track/Album</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SD</td>
<td>Steely Dan/Cousin Dupree / Two Against Nature</td>
<td>Pop/Jazz with male vocal</td>
</tr>
<tr>
<td>JW</td>
<td>Jennifer Warnes/Bird on a Wire / Blue Raincoat</td>
<td>Pop with female vocal</td>
</tr>
<tr>
<td>BSG</td>
<td>Stu Philips / Theme from Battle Star Galactica</td>
<td>Classical Orchestra</td>
</tr>
</tbody>
</table>

**ABSTRACT**

A listening test was conducted to identify music programs that provide sensitive, discriminating and reliable ratings for in-ear (IE) headphone evaluations. Ten trained listeners gave sound quality ratings for eight models of IE headphonest using ten different music programs. A virtualized headphone method was used to provide double blind, controlled presentations in which headphone leakage effects were monitored and eliminated. The main effect on the sound quality ratings was due to headphones while the program produced no significant effects or interactions. However, certain programs produced more discriminating and reliable ratings than other programs, the key factor being the bandwidth of the program’s spectral content, and the subject’s familiarity with it. As expected, the amount of bass content in each program tended to influence the ratings of headphones that had too much or too little bass output in their measured frequency response.
• These 3 programs were selected based on a pilot study [1] where listeners evaluated 8 headphones using 10 programs

• These three programs produced the most discriminating and reliable ratings among 10 programs

• No significant program effects or interactions

---

EXPERIMENTAL METHOD

- 5 listening tests completed in 2 blocks or sessions
- Multiple stimulus with 6 headphones, hidden reference and low anchor
- Presentation order of sessions/tests/headphones/programs randomized among subjects
- Headphones level matched for equal loudness (ITU-R BS 1770.3)
- Headphone leakage eliminated in subjective & objective measurements
- A virtual headphone method was used
Previous studies on AE (Temme et al, 2016) and IE headphones (Welti et al. 2016) found that in most cases the magnitude/minimum phase response accounts for most of differences in headphone preference ratings.
LISTENERS

- 71 listeners (Male = 82%, Female = 18%)
- Trained (n= 36)  Untrained (n= 35)
- Median age = 35 yrs, SD = 5 yrs
- Harman employees from Northridge, CA and Novi, Michigan
- All paid for service
- Trained listeners had normal audiometric hearing (untrained not tested)
RESULTS
### REPEATED MEASURES ANOVA SUMMARY

<table>
<thead>
<tr>
<th>Test</th>
<th>Headphone (HP)</th>
<th>Training</th>
<th>Program</th>
<th>HP * Trained</th>
<th>HP * Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>$F = 435.2, \ p &lt; 0.0001$</td>
<td>$F = 96.14; \ p &lt; 0.0001$</td>
<td>$F = 1.88; \ p = 0.153$</td>
<td>$F = 11.4; \ p &lt; .0001$</td>
<td>$F = 4.07; \ p &lt; 0.0001$</td>
</tr>
<tr>
<td>Two</td>
<td>$F = 271.53, \ p &lt; 0.0001$</td>
<td>$F = 80.11; \ p &lt; 0.0001$</td>
<td>$F = 2.62; \ p = 0.074$</td>
<td>$F = 11.2; \ p &lt; 0.0001$</td>
<td>$F = 4.18; \ p &lt; 0.0001$</td>
</tr>
<tr>
<td>Three</td>
<td>$F = 91.29, \ p &lt; 0.0001$</td>
<td>$F = 18.56; \ p &lt; 0.0001$</td>
<td>$F = 8.56; \ p = 0.648$</td>
<td>$F = 8.59; \ p &lt; 0.0001$</td>
<td>$F = 0.75; \ p=0.729$</td>
</tr>
<tr>
<td>Four</td>
<td>$F = 271.37, \ p &lt; 0.0001$</td>
<td>$F = 110.68; \ p &lt; 0.0001$</td>
<td>$F = 0.97; \ p = 0.38$</td>
<td>$F = 3.75; \ p &lt; 0.0001$</td>
<td>$F = 3.9; \ p &lt; 0.0001$</td>
</tr>
<tr>
<td>Five</td>
<td>$F = 408.46, \ p &lt; 0.0001$</td>
<td>$F = 86.94; \ p &lt; 0.0001$</td>
<td>$F =1.13; \ p = 0.325$</td>
<td>$F = 6.87; \ p &lt; 0.0001$</td>
<td>$F = 2.12; \ p = 0.009$</td>
</tr>
</tbody>
</table>

**Conclusions:**

1. Headphone was the dominant effect on listener preference, followed by Training
2. Training and Program had very small but statistically significant interactions with Headphone
TEST TWO

Preference Rating

Headphone Model

Target | HP7 | HP8 | HP9 | HP10 | HP11 | HP12 | Low Anchor

66.4 | 56.3 | 53.1 | 49.3 | 48.2 | 39.9 | 35.8 | 14.9
TEST FOUR

Preference Rating

Headphone Model

<table>
<thead>
<tr>
<th>Model</th>
<th>Preference Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>67.2</td>
</tr>
<tr>
<td>HP19</td>
<td>51.1</td>
</tr>
<tr>
<td>HP20</td>
<td>50.6</td>
</tr>
<tr>
<td>HP21</td>
<td>42.5</td>
</tr>
<tr>
<td>HP22</td>
<td>39.0</td>
</tr>
<tr>
<td>HP23</td>
<td>38.4</td>
</tr>
<tr>
<td>HP24</td>
<td>25.7</td>
</tr>
<tr>
<td>Low Anchor</td>
<td>18.6</td>
</tr>
</tbody>
</table>
**EFFECT OF LISTENER TRAINING ON PREFERENCE RATINGS**

**Conclusion:** Untrained listeners gave higher ratings (i.e. 10.4 points on average) than trained listeners. Consistent with prior research.
TRAINED VS UNTRAINED LISTENERS

**Test One**

- Preference Rating: 0 to 80
- Targets: HP1, HP2, HP3, HP4, HP5
- Anchor: HP6

Untrained vs Trained

**Test Two**

- Preference Rating: 0 to 80
- Targets: HP7, HP8, HP9, HP10, HP11, HP12
- Anchor: HP12

Untrained vs Trained
Test Five

Trained vs Untrained Listeners

Preference Rating

Untrained

Trained

Target
Target 2
HP25
HP26
HP27
HP28
HP29
Low Anchor
CONCLUSIONS ABOUT TRAINING EFFECTS

• Untrained listeners generally rated the headphones higher on the preference scale compared to trained listeners (consistent with prior research)

• Both trained and untrained listeners rated the Target highest and Low Anchor the lowest in general

• Ranking of the headphones was generally consistent across listening groups (some noted differences)

• Trained listeners were generally better at discriminating among the headphones in the middle of the scale (mediocre models)
OBJECTIVE
MEASUREMENTS
• The magnitude and minimum phase response of each headphone was measured with G.R.A.S. RA0045 (shown on right) externally polarized coupler equipped with a GR0408 nozzle according to IEC 60318-4

• This eliminated leakage effects

• Variations in headphone insertion depth were minimized using a standardized insertion force

• Measurements of all models are in the AES paper Part 1
The difference in level between the magnitude response of the Headphone and the Target curve

Only calculated to 10 kHz
Conclusion: Listeners tended to give lower preference ratings to headphones as their magnitude response deviated from the Target
• PLS was used to develop a linear predictive model to predict headphone preference ratings using statistical metrics based on deviations from the Harman IE Target.

• PLS is appropriate when the independent variables are collinear or there are a large number of variables compared to observations.

• PLS reduces the independent variables to a set of uncorrelated principal components.
INDEPENDENT VARIABLES FOR MODEL

\[ AS = \frac{\sum (\ln(x) - \ln(\bar{x}))(y - \bar{y})}{\sum (\ln(x) - \ln(\bar{x}))^2} \]

\[ ME = \frac{\sum_{i=1}^{n} \text{abs}(y_i)}{n} \]

\[ SD = \sqrt{\frac{\sum (y - \bar{y})^2}{n - 1}} \]

\( AS = \text{Absolute value of the Slope of the error curve (20 Hz to 10 kHz)} \)

\( ME = \text{Mean Error of the absolute deviation of the error curve from 40 Hz to 10 kHz} \)

\( SD = \text{the Standard Deviation of error in the error curve (20 Hz to 10 kHz)} \)
**GOODNESS OF FIT STATISTICS FOR MODEL**

<table>
<thead>
<tr>
<th>Goodness of Fit Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>32.000</td>
</tr>
<tr>
<td>Sum of weights</td>
<td>31.000</td>
</tr>
<tr>
<td>DF</td>
<td>29.000</td>
</tr>
<tr>
<td>R</td>
<td>0.91</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.819</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>5.691</td>
</tr>
<tr>
<td>MSE</td>
<td>30.301</td>
</tr>
<tr>
<td>RMSE</td>
<td>5.505</td>
</tr>
</tbody>
</table>

Predicted Preference = 68.685 - (3.238*SD) - (4.473 *AS) - (2.658 *ME)
• Listeners tend not use the top and bottom of the preference scale (contraction bias)

• This model is based on preference ratings transformed to fit the entire scale from 0 to 100

• Marketing prefers this model since the highest rated headphones are closer to 100%

Predicted Preference = 100.0795 - (8.50*SD) - (6.796*AS) - (3.475*ME)
Validation was done 2 ways using subsets of the 32 headphones tested and looking at goodness of fit statistics:

**Method One**

- Applying the statistical model to each of the five tests. How well does it predict the actual preference ratings?

**Method Two (Cross-validation)**

- Building new models by randomly removing 6 or 10 headphones from the sample (repeated 10 times)
## VALIDATION TEST ONE: GOODNESS OF FIT

<table>
<thead>
<tr>
<th>Test</th>
<th>Correlation Coefficient ($r$)</th>
<th>Error (RMSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>0.96</td>
<td>4.95</td>
</tr>
<tr>
<td>Two</td>
<td>0.95</td>
<td>5.58</td>
</tr>
<tr>
<td>Three</td>
<td>0.89</td>
<td>6.35</td>
</tr>
<tr>
<td>Four</td>
<td>0.90</td>
<td>5.46</td>
</tr>
<tr>
<td>Five</td>
<td>0.98</td>
<td>3.81</td>
</tr>
</tbody>
</table>

**Conclusion:** The model produced accurate predictions of headphone preference ratings for Tests One through Five, indicating it is robust.
### VALIDATION TEST TWO: GOODNESS OF FIT

<table>
<thead>
<tr>
<th>Number of headphones removed</th>
<th>Range of r values *</th>
<th>Mean r value</th>
<th>Range of error values</th>
<th>Mean Error value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.87 to 0.96</td>
<td>0.91</td>
<td>4.4 to 6.1%</td>
<td>5.5%</td>
</tr>
<tr>
<td>10</td>
<td>0.83 to 0.95</td>
<td>0.90</td>
<td>4.4 to 6.1%</td>
<td>5.5%</td>
</tr>
</tbody>
</table>

* R and Error values based on a repetition of 10 tests

**Conclusion:** The model produced good predictions of preference independent of the selection and number of headphones indicating it is robust and not over-fitted.
RETAIL PRICE VERSUS SOUND QUALITY

(32 different models of IE headphones from 19 manufacturers)

Conclusion: There is very little correlation between price and sound quality

$r = 0.138$
CONCLUSIONS

• Harman IE Target curve preferred to 30 other models for both trained and untrained listeners

• Untrained listeners gave higher ratings and were less discriminating; otherwise similar headphone preferences to trained listeners

• New statistical model predicts headphone ratings with correlation of \( r = 0.91 \) with 5.5% error based on magnitude deviations from Harman IE Target response

• Validation of the model shows it’s robust and not over-fitted

• Needs to be further tested with new sample of headphones
THE PERCEPTION AND MEASUREMENT OF AROUND-EAR / ON-EAR HEADPHONES (WORK IN PROGRESS)

SEAN OLIVE
TODD WELTI
OMID KHONSARIPOUR

July 14 2017
Over 30 models covering from different manufacturers covering a price range of $80 - $4k are being tested using 200+ listeners both trained and untrained.
LISTENING TESTS TO DATE

• 126 listeners participated at 3 Harman locations in Northridge (12), Karlsbad/Garching Germany

• both trained and trained

• Additional tests are being done in California, China and Michigan

• Also looking into how context affects headphone preference
CONTEXT EFFECT

Target

Low Anchor

Headphone A

Headphone B

Headphone O

Headphone P

Headphone V

All Listeners

Mean

Minimum/Maximum

Outliers(1)

Outliers(2)
EXAMPLE OF CONTEXT EFFECT

TEST ONE (MANY DULL HPs)

TEST TWO (FEWER DULL HEADPHONES)
SUMMARY

- What is the optimal headphone response in terms of sound quality?
  - YES! but varies with program and to some extent listener age/experience
- Is it the same for Around Ear (AE) / On Ear (OE) and In-ear (IE) headphones?
  - NO! Listeners seem to prefer 4-6 dB more bass in IE headphones
- To what extent do listeners agree on the preferred target response?
  - Good agreement notwithstanding age/hearing loss/experience
- Can we predict listeners’ headphone sound quality preferences based on objective measurements?
  - YES !! 91% accuracy for IE headphones and probably similar results with AE/OE if leakage effects are controlled
CONCLUSIONS

• There is a scientific basis for choosing loudspeakers and headphones

• Listeners prefer models that are accurate and neutral across age, listening experience, or culture with some slight bass/treble variations to account for program/age/training/hearing loss

• Perceived sound quality can be accurately predicted based on objective measurements (sadly these measurements are not readily available from manufacturers and not part of current international standards like ITU and EBU)

• To improve the consistency and quality of recordings and reproduction we must have loudspeakers and headphones that are similarly uniform throughout the recording/reproduction chain
Microphones & recording processes are judged by listening through.....

.....to judge the quality of loudspeakers and headphones in product development and by consumers.....

.....by listening to recordings used to..

SOLVING AUDIO’S “CIRCLE OF CONFUSION”
BREAKING THE CIRCLE OF CONFUSION

JBL Pro M2

Revel Performa F208
BREAKING THE CIRCLE OF CONFUSION

JBL Pro M2

Revel Performa F208

CREATION OF THE ART

PRESERVATION OF THE ART
BREAKING THE CIRCLE OF CONFUSION

JBL Pro M2

JBL Everest 700 Elite
BREAKING THE CIRCLE OF CONFUSION

JBL Pro M2

JBL Everest 700 Elite

CREATION OF THE ART

PRESERVATION OF THE ART
Without a meaningful loudspeaker and headphone standard shared by both professional and consumer audio industries...

there is confusion and uncertainty in the sound quality throughout the production and reproduction of sound
With a meaningful loudspeaker and headphone standard shared by both professional and consumer audio industries...

We can produce and reproduce consistent high quality content with CONFIDENCE.