

NASA/TM-20250008616



# Sound Level Calibration of a Kiosk for Psychoacoustic Testing

*Aaron B. Vaughn  
Langley Research Center, Hampton, Virginia*

---

September 2025

## NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.

- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at <http://www.sti.nasa.gov>
- E-mail your question to [help@sti.nasa.gov](mailto:help@sti.nasa.gov)
- Phone the NASA STI Information Desk at 757-864-9658
- Write to:  
NASA STI Information Desk  
Mail Stop 148  
NASA Langley Research Center  
Hampton, VA 23681-2199

NASA/TM-20250008616



# Sound Level Calibration of a Kiosk for Psychoacoustic Testing

*Aaron B. Vaughn  
Langley Research Center, Hampton, Virginia*

National Aeronautics and  
Space Administration

Langley Research Center  
Hampton, Virginia 23681-2199

---

September 2025

The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

Available from:

NASA STI Program / Mail Stop 148  
NASA Langley Research Center  
Hampton, VA 23681-2199  
Fax: 757-864-6500

This page intentionally left blank.

<b>Table of Contents</b>	<b>Page</b>
Nomenclature and Acronyms .....	vii
Executive Summary .....	viii
I. INTRODUCTION .....	1
II. METHODOLOGY.....	1
A. Psychoacoustic Testing Kiosk .....	1
B. Kiosk Calibration Setup .....	2
C. Acoustic Data Collection .....	2
III. RESULTS .....	4
A. Spectral Response .....	4
B. OASPL vs Volume.....	6
C. VANGARD Kiosk Test Calibration .....	7
D. Measurement Variability.....	11
1. Headphone Donning .....	11
2. Calibrated Volume Settings .....	13
IV. CONCLUDING REMARKS.....	14
References.....	15
Acknowledgments.....	16
Appendix A : VANGARD Kiosk Test Calibration Stimuli.....	16

## Nomenclature and Acronyms

AAM	Advanced Air Mobility
ASEL	A-weighted sound exposure level
DAC	digital-to-analog converter
dB	decibel
HME	HEAD Microphone Electronic
HMM	HEAD Measurement Microphone
HMS	Head Measurement System
Hz	hertz
kHz	kilohertz
NI	National Instruments
$LA_{max}$	A-weighted maximum sound level
MATLAB	MATrix LABoratory
OASPL	overall sound pressure level
OTO	one-third octave
PCI	peripheral component interconnect
PSD	power spectral density
PXI	PCI eXtensions for Instrumentation
SPL	sound pressure level
USB	universal serial bus
VANGARD	Varied AAM Noise and Geographic Area Response Difference

## **Executive Summary**

This document details the calibration process of kiosks for psychoacoustic testing. NASA developed a remote psychoacoustic testing capability to enable participation of geographically diverse members of the public using their personal audio equipment. This capability is used in the Varied Advanced Air Mobility (AAM) Noise and Geographic Area Response Difference (VANGARD) test. When people use different audio equipment, it creates uncertainty in test results. Calibrated kiosks with known playback levels can be used with a small control group of test subjects to address this uncertainty. The current kiosk design consists of a Microsoft Surface Go 3 tablet computer and Sennheiser HD 280 headphones. Two options are available for digital-to-analog conversion of sound stimuli: the sound card in the Surface Go 3 computer or a DragonFly USB digital-to-analog converter (DAC) unit. The kiosk calibration process described here is performed using a manikin head with in-ear microphones. A pink noise signal is used to determine the spectral response, sound level as a function of volume setting, and measurement variability. VANGARD test stimuli are played through the kiosk and measured with the manikin head to determine the tablet computer volume setting needed to play back stimuli at their intended sound level. The calibration process determined that for the VANGARD test a volume setting of 47 is appropriate when using the kiosk with the onboard sound card and 21 with the DragonFly DAC. These settings ensure that most sounds play within  $\pm 2$  dB of the intended A-weighted sound exposure level (ASEL) and maximum sound level ( $LA_{max}$ ).

## I. INTRODUCTION

NASA has developed a remote psychoacoustic testing capability (Krishnamurthy et al., 2024). This approach enables testing with a much wider population but introduces uncertainties from varying personal audio equipment and self-calibration methods. To quantify these uncertainties, test kiosks have been developed that use well-characterized audio equipment for a small control group. By comparing responses between this control group and the larger remote test population, researchers can better understand how personal audio equipment affects test results. These kiosks require precise calibration to ensure all participants hear sounds at the intended levels.

This document describes a procedure for calibrating kiosks by determining the tablet volume setting that produces accurate sound levels at the participant's ear. The procedure examines the acoustic response of the kiosk setup and determines a volume setting for use with the Varied Advanced Air Mobility (AAM) Noise and Geographic Area Response Difference (VANGARD) test (NASA, 2024). This calibration method can also be used for future psychoacoustic tests.

## II. METHODOLOGY

This section details the kiosk calibration methodology. The kiosk equipment is described, followed by descriptions of the recording equipment, test signals, and acoustic recording process.

### A. Psychoacoustic Testing Kiosk

NASA procured 20 kiosks for psychoacoustic testing. Each kiosk includes a Microsoft Surface Go 3 tablet computer and a pair of Sennheiser HD 280 headphones, as shown in Figure 1a. The headphones feature a closed, circumaural (over-the-ear) design. These headphones cover the entire audible human frequency range with specifications of 8 Hz to 25 kHz (Sennheiser, 2025). Pickens et al. (2018) demonstrated that these headphones maintain a flat frequency response from 100 Hz to 10 kHz in their study evaluating headphones for hearing screening.

NASA procured four DragonFly Cobalt USB digital-to-analog converter (DAC) units, or simply DragonFly DACs, for laboratory testing. The DragonFly DAC is shown in Figure 1b. The DragonFly DAC is a 24-bit audio interface that provides an alternative to the built-in DAC of the computer. The DragonFly DAC is used with the kiosk to reduce sound level variance, similar to other psychoacoustic studies (Sørensen et al., 2023; Torija Martinez and Li, 2020; Gee et al., 2018). The DragonFly DAC is intended for in-person tests conducted in the laboratory at NASA Langley Research Center, as a USB DAC may be unfamiliar to participants and may be easily misplaced due to its small size.



Figure 1. a) Microsoft Surface Go 3 tablet computer and a pair of Sennheiser HD 280 headphones, b) DragonFly Cobalt USB DAC, and c) HEAD Measurement System with the psychoacoustic testing kiosk in a small anechoic chamber.

## B. Kiosk Calibration Setup

A HEAD Acoustics artificial head measurement system (HMS) measured the audio output from the kiosk through the headphones for calibration. The HMS and other manikin heads for acoustic research are generally used to imitate a human listener for headphone evaluations (Kinnunen, 2022; Hirvonen et al., 2003), headphone calibrations (Ooi et al., 2021; Xie et al., 2009), development of virtual acoustic environments (Whiting, 2018; Johansson et al., 2012), and experimental validation of acoustical theories (Bellows, 2023). The HMS consists of a HEAD Measurement Microphone (HMM) II.21 and a HEAD Microphone Electronics (HME) II.21, as shown in Figure 1c and Figure 2. The HMM II.21 consists of a manikin head with microphones mounted inside the ears to simulate measurements of a human listener for binaural recordings. Notably, the HMM II.21 has no ear canals, which has implications for the frequency response of the kiosk calibration setup as noted in Section III.A. The HME II.21 is inside the torso of the HMS and houses the input/output ports, internal calibrator, and power supply.

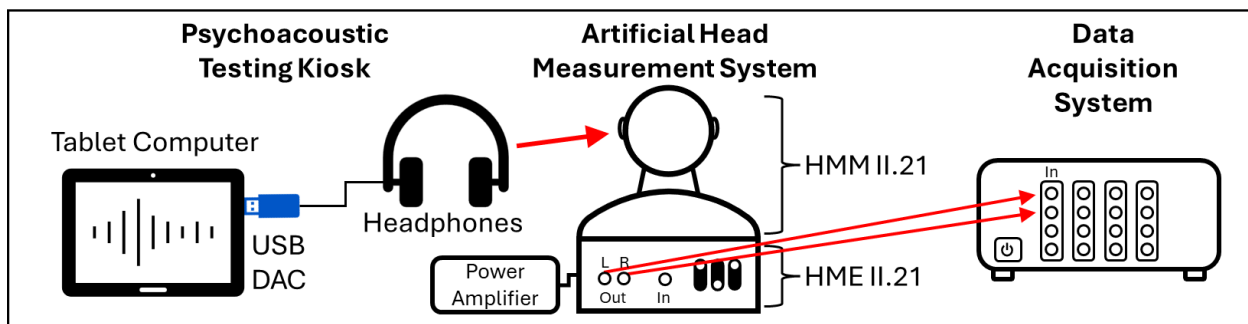


Figure 2. Kiosk calibration setup schematic (red arrows show signal path).

A schematic of the kiosk calibration setup is depicted in Figure 2 with red arrows showing the signal path between systems. In conjunction with the HMS, National Instruments (NI) hardware (NI PXI-1042 chassis and an NI PXI-4462 card) and m+p Analyzer software were used to obtain acoustic data as part of the Data Acquisition System. A sampling rate of 51.2 kHz was utilized to set the Nyquist frequency above the upper limit of human hearing at 20 kHz. The in-ear microphones are 1/2-inch free-field Microtech Gefell MK 221 microphones with a dynamic range of 15 to 149 dB.

## C. Acoustic Data Collection

Acoustic data for kiosk calibration were collected in a small anechoic chamber (Room 145 of Building 1208 at NASA Langley Research Center), as pictured in Figure 1c. This location served as a quiet space to minimize external or ambient noise effects on the recordings. Data were collected from two channels, one for each ear. Measurements made with the DragonFly DAC inform the tablet computer volume setting for tests conducted in the laboratory at NASA Langley Research Center, such as the VANGARD kiosk test, while measurements made with the onboard sound card are useful for calibrating kiosks that are shipped to subjects in remote tests where the DAC onboard the tablet computer is used.

Data collection used pink noise as the input signal. Pink noise produces a known flat spectrum across one-third octave (OTO) bands, which avoids the need to measure the input spectra for kiosk spectral response analysis. The pink noise signal was generated and played back over the tablet computer using Audacity® (version 3.5.1). The signal was 1-minute long, and the digital gains were adjusted to ensure adequate signal-to-noise ratio and to avoid signal clipping by the DAC, with the final amplitude set to 0.3 in the software.

Recordings were then made of the pink noise after it was output by the headphones. With the DragonFly DAC inserted and no adjustment made to the headphones between recordings, 30-second recordings were made at tablet computer volume settings from 5 to 100 in increments of 5. The same procedure was used to characterize the output of the Surface Pro onboard DAC.

Using stimuli from a particular test may provide a more meaningful calibration, as frequency content will likely differ from pink noise and may produce slightly different intended sound levels. To determine the appropriate tablet computer volume setting for the kiosk portion of the VANGARD test, measurements were made using stimuli prepared for the VANGARD test. Table 1 lists the flight condition; monoaural (or pre-binauralized) A-weighted sound exposure level (ASEL) and maximum sound level ( $LA_{max}$ ); and signal length of four preliminary VANGARD test signals. The signals were full aircraft flyovers. The ASEL is evaluated over the 10-dB down region about the maximum level and represents the main contribution of acoustic exposure for the noise event, while the  $LA_{max}$  indicates the maximum level occurring in the noise event. The test signals encompass the range of levels and flight conditions representative of the stimuli for the VANGARD test. To observe the general trend of level versus volume, the four signals were measured at various volume settings in increments of 10 from 100 to 10. The results were used to identify a range of volume settings where the output ASEL closely matched the intended ASEL values from Table 1. Discrete measurements in increments of 1 were then made to determine the proposed volume setting. Finally, measurements were made for 43 stimuli from the VANGARD test at the proposed volume setting to characterize the measurement variability.

Table 1. Signals from the VANGARD test stimuli for volume calibration.

Signal Number	Flight Condition	Monoaural ASEL (dB)	Monoaural $LA_{max}$ (dB)	Signal Length (s)
1	Approach	74.0	64.3	30
2	Cruise	55.3	48.5	30
3	Departure	86.5	76.7	30
4	Approach	63.2	54.6	20

To further examine measurement variability, additional measurements were made of pink noise with various headphone configurations shown in Figure 3. Participants in VANGARD and other headphone-based psychoacoustic tests will be instructed on the proper wearing of the headphones; however, each individual has a different combination of head, hair, and ear shape that makes uniform donning of the headphones difficult. Therefore, the variability due to headphone donning can be examined by measuring extreme headphone configurations. Relative to the regular configuration, the headphones were flipped for the “Reverse” configuration; the strap was fully tightened for the “Tight” configuration and fully loosened for the “Loose” configuration; and the headphones were tilted forward for the “Forward” configuration and backward for the “Backward” configuration. Between each measurement, the headphones were removed and replaced on the head.



Figure 3. Six headphone configurations used to examine measurement variability.

### III. RESULTS

Pressure power spectral density (PSD) results are shown to depict the spectral response followed by overall sound pressure level (OASPL) versus volume setting for the pink noise signal. Kiosk calibration measurements are then provided for the VANGARD test. Measurement variability due to headphone donning is then shown at proposed volume settings.

#### A. Spectral Response

The kiosk spectral response is examined for consistency across volume settings. Figure 4 shows the PSD response to pink noise at OTO frequencies for the kiosk setup with the DragonFly DAC and with the onboard sound card, measured for the left and right channels. The volume setting is varied from 5 to 100 in increments of 5. A 10-dB offset is the main difference in spectral response between setups with the DragonFly DAC and the onboard sound card. This offset is observed throughout the analyses and indicates the volume setting should be lower when the onboard sound card is used than when the DragonFly DAC is used.

The spectral shapes in Figure 4 share similar characteristics for both left/right channels and both the DragonFly DAC and onboard sound card. The overall spectral response is relatively flat, especially at mid-range frequencies from 200 to 2,000 Hz. The response at lower frequencies (< 200 Hz) has more variability than the mid-range frequencies. differs between the left and right channel. There is a dip at 3 to 4 kHz, which corresponds to the typical resonance frequency of the human ear canal ([Strong and Plitnik, 2007](#)). The headphones are designed to compensate for this resonance, and since the HMM II.21 has no ear canal, a dip is observed in the spectral response. The response then rolls off above 8 kHz.<sup>1</sup>

Spectral shape consistency across volume settings is depicted in Figure 5. Volume settings below 15 (i.e., 5 and 10) are excluded as they are too quiet to produce intended levels for the VANGARD test and were more variable, possibly due to noise floor contamination. As a check of linearity, the spectrum for each volume setting is normalized by the level at 1 kHz in Figure 5a and 4b. This process collapses the spectra across the multiple volume settings and reveals a consistent spectral shape across all volume settings for a given channel. The maximum level difference across the normalized spectra is given in Figure 5c and 4d as a function of OTO frequency for each channel. For both channels and setups, the maximum level differences are less than 1 dB from 40 Hz to nearly 20 kHz and less than 0.2 dB from 200 Hz to 10 kHz, suggesting minimal to no distortion across volume settings. These results indicate that for volume settings greater than 10, one can expect approximately the same spectral response regardless of the selected volume setting.

<sup>1</sup> [Pickens et al. \(2018\)](#) noted that the Sennheiser HD280 headphones had decreased detection probability at 4 kHz and 8 kHz in their study.

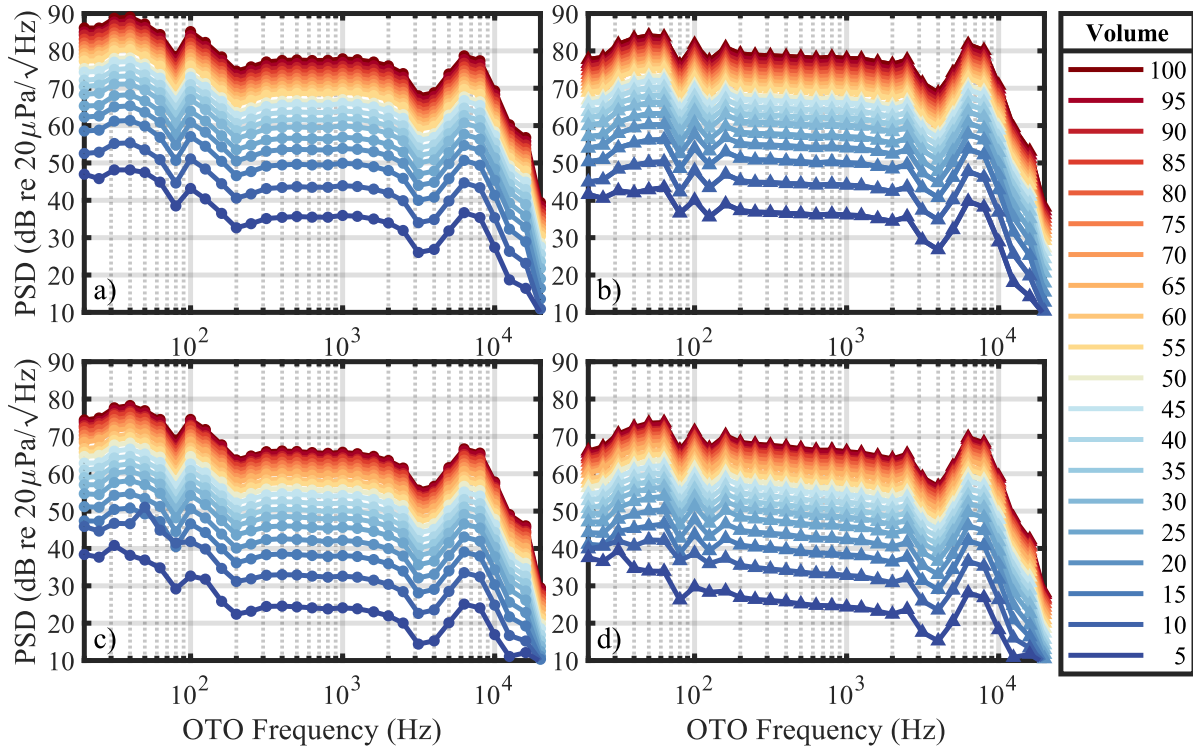


Figure 4. PSD response to pink noise in OTO frequencies for a) left and b) right channels with the DragonFly DAC and c) left and d) right channels with the onboard sound card.

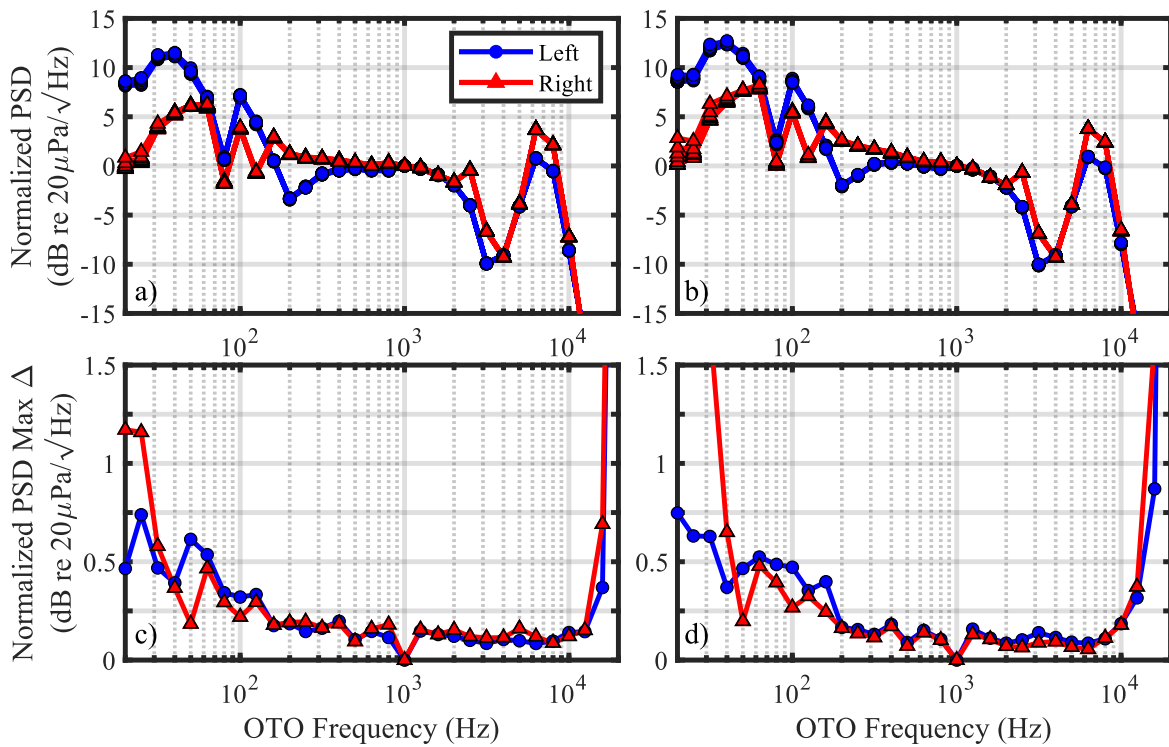


Figure 5. Normalized kiosk spectral response of pink noise in OTO frequencies a) with the DragonFly DAC and b) with the onboard sound card and the maximum differences across volume settings of the normalized spectra c) with the DragonFly DAC and d) with the onboard sound card.

While the spectral shapes are similar across volume settings for a given channel, the spectral shape differs between the left and right channels. Figure 6 depicts the kiosk spectral response with the volume set to 50. The OASPL differs between channels by about 3 dB. The spectral shapes differ most at lower frequencies, which have the highest sound level. This low-frequency difference is likely due to how the headphones are donned and the seal about the muff and the ear (e.g., see [DIY Audio Heaven, 2023](#)). A “good” seal has more low-frequency content and a “poor” or broken seal allows for leakage of low-frequency content ([Poldy, 2001](#)). Therefore, the left channel with higher low-frequency levels appears to have a better seal for both setups with the DragonFly DAC and with the onboard sound card. Given that the headphones were not adjusted between measurements with the DragonFly DAC and with the onboard sound card, similar trends between the two setups are expected.

Applying a frequency weighting that accounts for the response of the human ear decreases the influence of low-frequency variability in the measured data. A-weighting, which deemphasizes the lower frequencies, is commonly applied to noise metrics used for assessing vehicle noise ([WHO, 2018](#); [Schultz, 1978](#)), including AAM vehicles ([Rizzi et al., 2020](#)). A-weighted spectra are shown in Figure 6 relative to the unweighted spectra for each channel. The difference in OASPL between channels is reduced for both setups to about 1.4 dB. These results indicate that while differences between channels may exist for a given measurement, OASPL differences are reduced when A-weighting is applied.

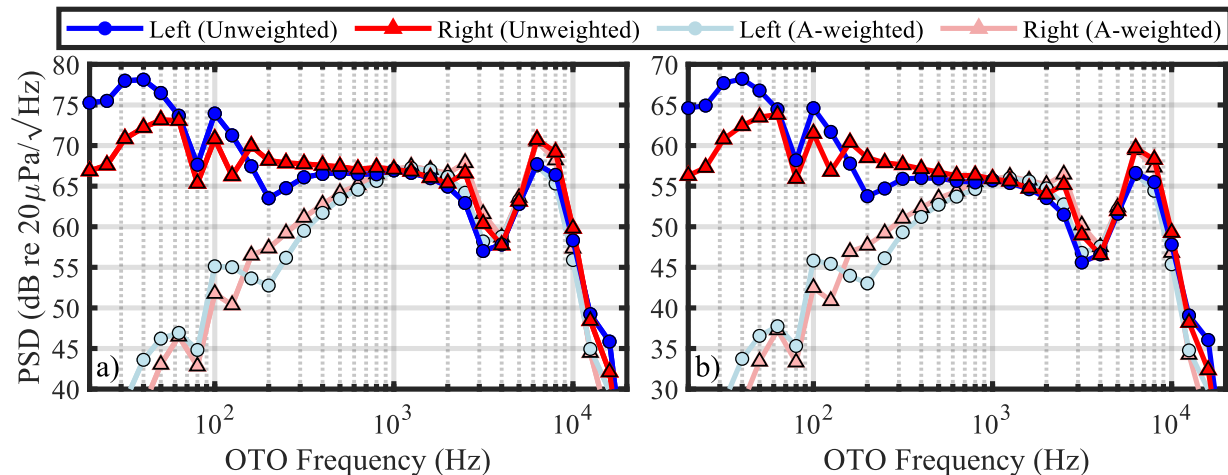


Figure 6. Kiosk spectral response of pink noise in one-third octave center band frequencies at a volume setting of 50 for the setup a) with the DragonFly DAC and b) with the onboard sound card.

## B. OASPL vs Volume

To determine the tablet computer volume setting for the VANGARD kiosk test, the OASPL of the pink noise signal is examined as a function of volume setting. The desired kiosk test volume setting is the one that achieves the same OASPL as an intended test sound. The OASPL is averaged over 30 seconds of pink noise and the volume setting is varied from 5 to 100 in increments of 5. Unweighted and A-weighted OASPL results are shown in Figure 7 for both channels and setups with the DragonFly DAC and with the onboard sound card.

There is a level offset between the channels for both setups and frequency weightings. For the unweighted OASPL with both setups, the left channel is greater than the right by about 3 dB at higher volume settings. The difference decreases at lower volume settings. This offset is mainly attributed to the previously discussed low-frequency differences due to headphone donning.

For the A-weighted OASPL, the trend flips for both setups, with the right channel being greater than the left by about 1.4 dB. This offset is consistent across all volume settings. Because A-weighting diminishes the contribution of low frequencies to the OASPL, differences in higher frequencies are the likely source of this offset. Examination of the spectra in Figure 6 reveals a level difference from 6 to 8 kHz. The physical phenomenon causing this difference is unknown.

Despite the differences between channels, the OASPL versus volume setting trends appear consistent and can help determine a volume setting. For instance, a functional fit could be used to interpolate across volume settings from 5 to 100. Interestingly, the trend for both frequency weightings with the DragonFly DAC in Figure 7a is exactly linear for volume settings  $\geq 55$ . As an example of determining a volume setting, if the pink noise signal was the intended sound waveform in units of pascals, then the target A-weighted OASPL is 66.3 dB. This suggests a calibrated volume setting of about 20 with the DragonFly DAC and 45 with the onboard sound card. Measurement variability discussed in Section III.D helps provide confidence bounds on a selected volume setting. Test stimuli from a particular likely has a spectral shape that differs from pink noise; therefore, the calibration process should include test stimuli.

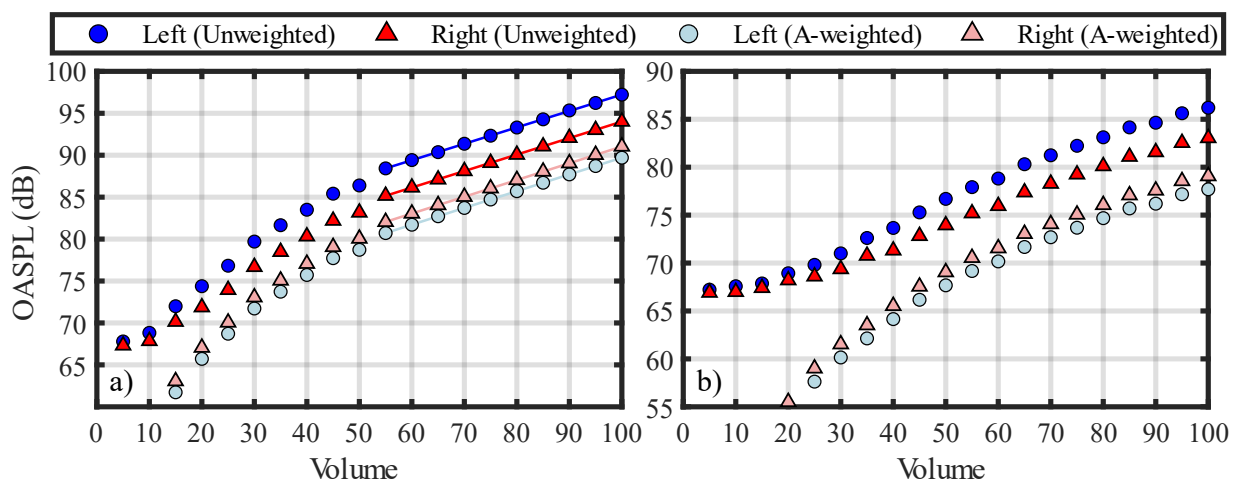


Figure 7. OASPL versus volume setting using pink noise for the setup a) with the DragonFly DAC and b) with the onboard sound card.

### C. VANGARD Kiosk Test Calibration

To calibrate the tablet computer volume setting for the VANGARD kiosk test, signals noted in Table 1 were first measured in volume increments of 10. These increments show the general trend across volume and help determine where to make additional discrete measurements to determine the appropriate volume setting. For Signal 1 (Approach AAM vehicle flight with 74 dB monaural ASEL), Figure 8 shows the measured 1-second block linear averaged (slow-weighting) A-weighted SPL calculated using a time domain filter bank. The test signals are binaural, meaning that head scattering effects are included through a head-related transfer function and the levels vary between left and right channels to simulate the sound source in motion. The AAM vehicle travels right to left for Signal 1, as noted by higher levels at the start of the signal in the right channel (see Figure 8b and 8d) relative to the left channel (see Figure 8a and 8c) and then higher levels at the end of the signal in the left than right channel. The SPL shape is similar across measurements but differs slightly from the intended shape. For instance, the intended signal in Figure 8d matches the measurement at a volume of 50 from about 0 to 15 s but is closer to a volume of 40 from about 24 to 30 s. Based on the results in Figure 8 for Signal 1, the calibrated volume

setting is expected to be a little greater than 20 with the DragonFly DAC and slightly less than 50 with the onboard sound card. Results for Signals 2, 3, and 4 are shown in Figure 17, 18, and 19 in Appendix A.

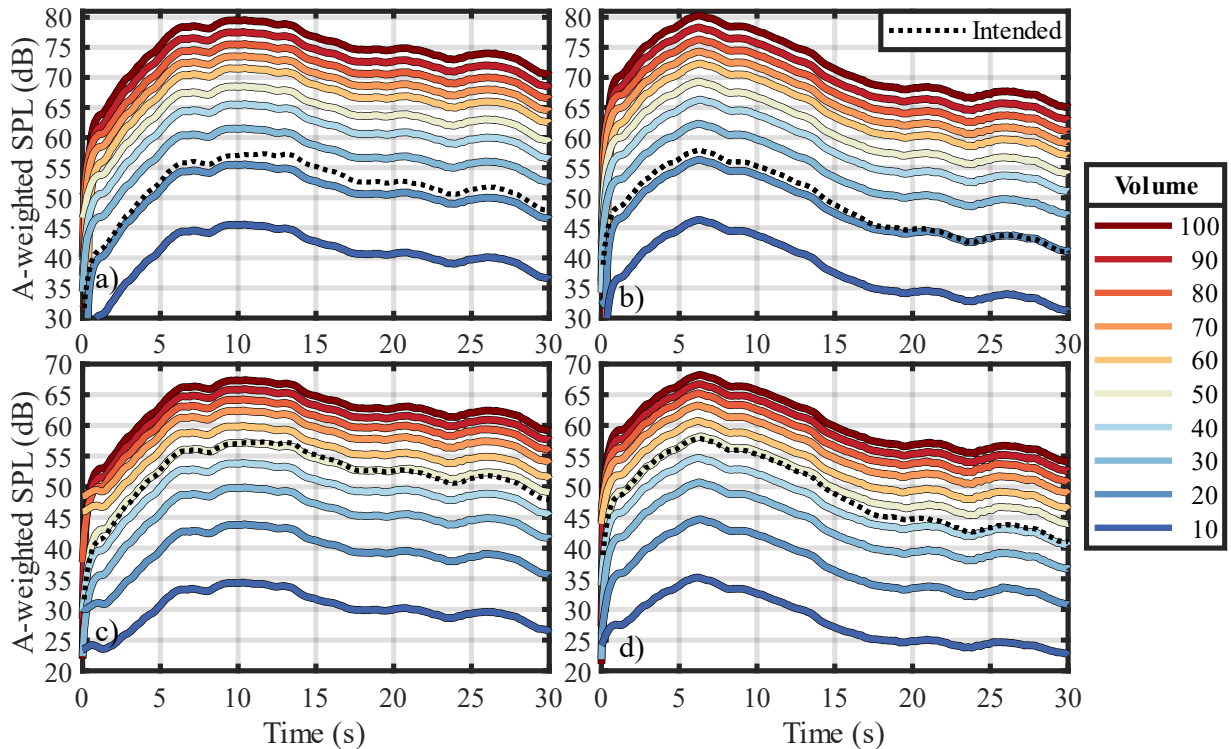


Figure 8. VANGARD test Signal 1 measured versus intended slow-weighted, A-weighted SPL versus time for the a) left and b) right channels with the DragonFly DAC and c) left and d) right channels with the onboard sound card.

Comparisons of measured ASEL relative to intended level for the four VANGARD test signals from Table 1 are shown in Figure 9 and 10. The ASEL is evaluated over the 10-dB down region about the maximum level and represents the main contribution of acoustic exposure for the event. For example, the ASEL for the right channel of Signal 1 in Figure 8d is evaluated over the first 17 s because the maximum level is at about 7 s and the 10-dB down covers roughly the first 17 s of the signal. The measurements of ASEL in Figure 9 and 10 are the gain offsets relative to the intended signal level to simply compare results across the four signals. Measurements of ASEL gain for the left and right channels are first shown in volume increments of 10 in Figure 9a and 8b with the DragonFly DAC and in Figure 11a and 11b with the onboard sound card. These plots show the general trend of ASEL increasing with volume.

Results at finer discrete volume settings are shown in Figure 9c and 9d and Figure 11c and 11d. Results of measurements in volume increments of 10 informed where measurements were made in discrete volume increments of 1. These discrete measurements help determine the appropriate volume settings to play back intended sound levels with minimal variation. Comparing the ASEL between channels, a given signal is 0.1 to 0.7 dB greater in the right channel relative to the left. The discrete results also reveal that the volume setting is discretized and does not continuously vary with level (i.e., the volume pairs 21 and 22, 45 and 46, and 48 and 49 produce

the same levels within each pair). Similar results are given for the slow-, A-weighted maximum sound level, or  $LA_{max}$ , in Figure 20 and 21 in Appendix A.

There is currently no mechanism to automatically and adaptively change the volume setting for each test sound with the kiosk tablet computers. A single volume level will need to be set before the test. This selected VANGARD kiosk test volume level should minimize, over all test sounds, the differences between intended ASEL and the ASEL achieved with the volume setting.

A tablet computer volume setting of 21 (or equivalently 22) appears appropriate for the VANGARD test with the DragonFly DAC. Figure 10a shows the average ASEL gain across the four VANGARD test signals. For both left and right channels in Figure 10a, a volume settings of 21 (and 22) is closest to an ASEL gain of 0 dB and contains 0 dB within  $\pm 1$  standard deviation. Similar to the ASEL results, the  $LA_{max}$  results in Figure 10b also support a volume setting of 21. This volume setting is agreeable with the A-weighted OASPL pink noise example from Section III.B. Additionally, the A-weighted SPL results for Signal 1 shown in Figure 8 seem supportive of a volume of 21, though ASEL results in Figure 9c and 9d indicate a volume of 23 is a better fit for Signal 1. The reason for differing results between the four signals is not readily apparent but may be due to the varying frequency content between the signals.

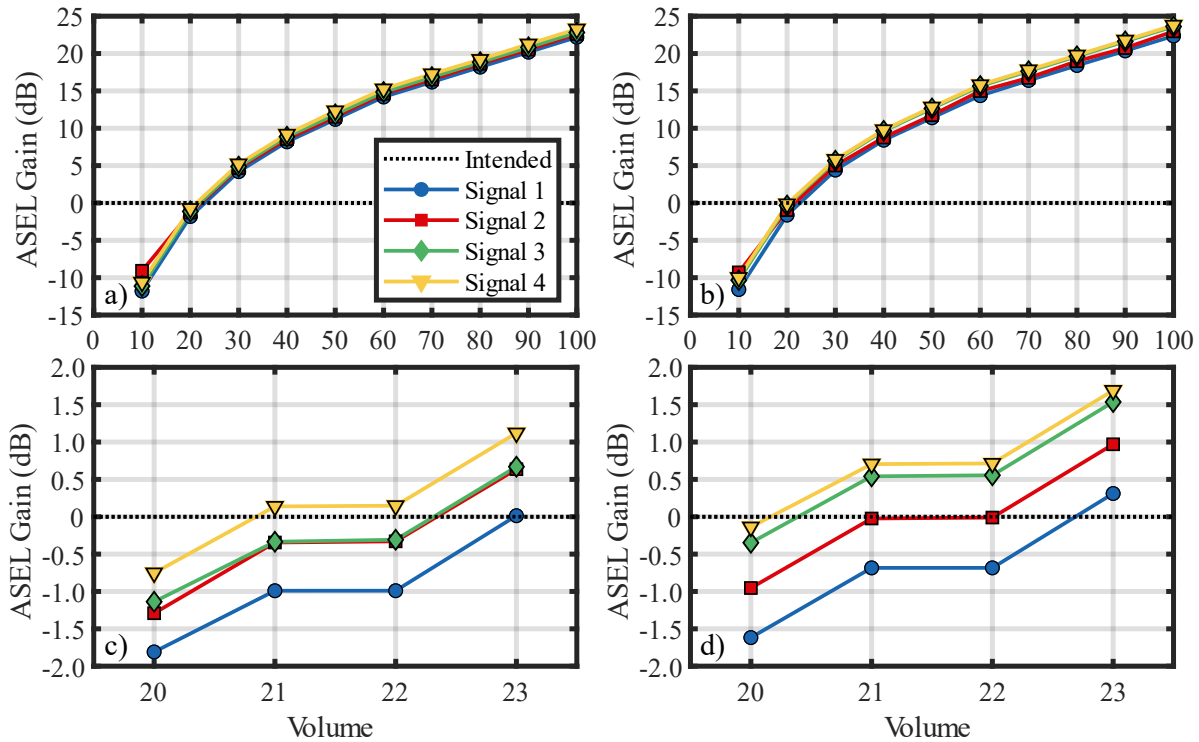


Figure 9. Gain of ASEL measurements with the DragonFly DAC relative to the intended level in volume increments of 10 for the a) left and b) right channels and in volume increments of 1 near ASEL gain of 0 dB for the c) left and d) right channels.

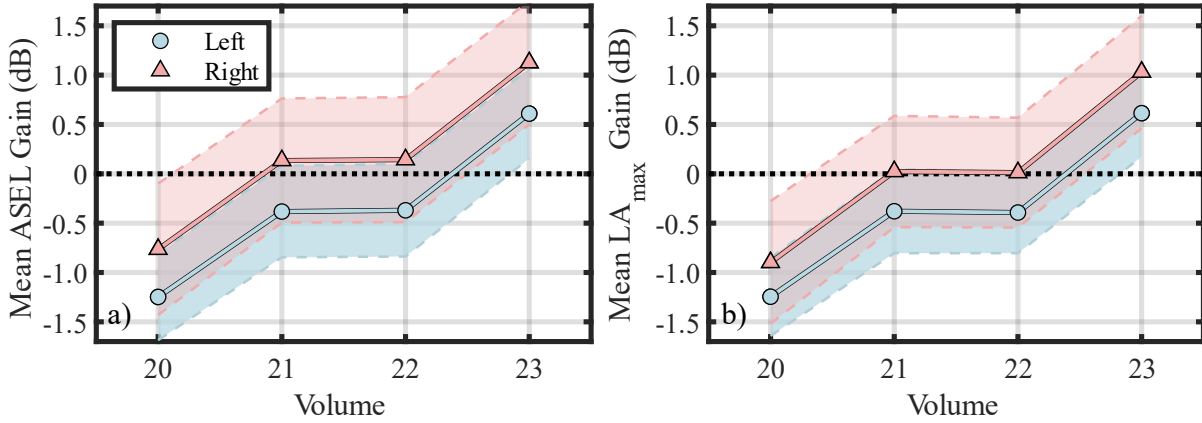


Figure 10. Average gain of a) ASEL and b) LA<sub>max</sub> measurements relative to the intended level for VANGARD test signals 1 to 4 with the DragonFly DAC. Shaded regions indicate  $\pm 1$  standard deviation.

A volume setting of 47 seems suitable for the VANGARD test with the onboard sound card. For both left and right channels in Figure 12a, volumes 45 to 47 are closest to an ASEL gain of 0 dB and contain 0 dB within  $\pm 1$  standard deviation. Measurements with the onboard sound card have larger standard deviation values, as noted in Table 2. The LA<sub>max</sub> results in Figure 12b support a volume setting of 47 because the average LA<sub>max</sub> and confidence bounds are more centered about 0 dB. A volume setting of 47 is also agreeable with the A-weighted OASPL pink noise example from Section III.B.

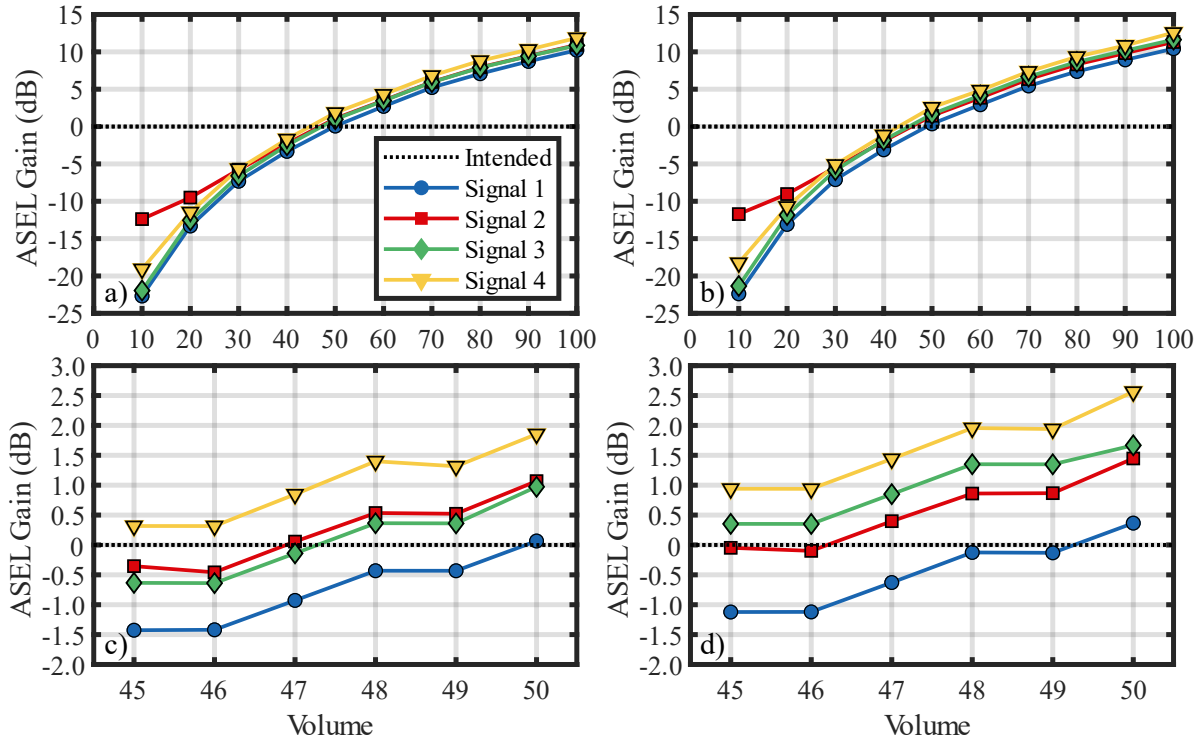


Figure 11. Gain of ASEL measurements with the onboard sound card relative to the intended level in volume increments of 10 for the a) left and b) right channels and in volume increments of 1 near ASEL gain of 0 dB for the c) left and d) right channels.

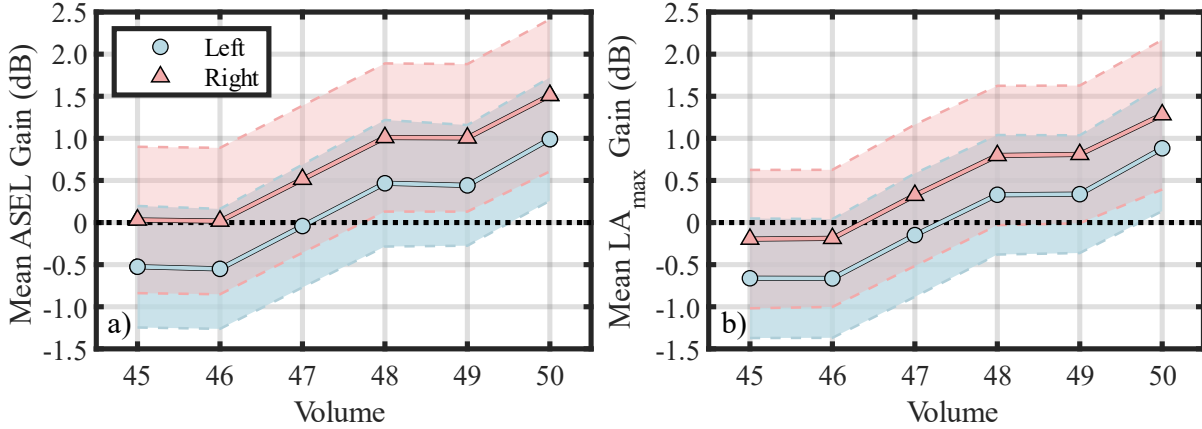


Figure 12. Average gain of a) ASEL and b) LA<sub>max</sub> measurements relative to the intended level for VANGARD test signals 1 to 4 with the onboard sound card. Shaded regions note the  $\pm 1$  standard deviation.

Table 2. Average standard deviation across measurements of the four VANGARD test signals.

Setup	ASEL (dB)		LA <sub>max</sub> (dB)	
	Left	Right	Left	Right
DragonFly DAC	0.4	0.6	0.4	0.6
Onboard Sound Card	0.7	0.9	0.7	0.8

#### D. Measurement Variability

Additional measurements were made to characterize measurement variability. First, measurements made at various headphone configurations with pink noise explored how levels may vary from person to person due to headphone donning. Then, measurements made with the VANGARD test signals at the calibrated volume settings were examined to provide confidence bounds on the calibrated volume settings.

##### 1. Headphone Donning

Additional measurements of pink noise were made at six headphone configurations (see Figure 3) to characterize variability due to headphone donning. Figure 13 shows the unweighted OASPL results. Three measurements at each configuration were made with the DragonFly DAC setup shown in Figure 13a and two measurements at each configuration with the onboard sound card shown in Figure 13b. These few measurements provide a glance into the variability due to headphone donning. The volume was set to 50 for these measurements. Given the same spectral shape across volume settings shown in Figure 5, trends at a volume of 50 are expected to hold for other volume settings.

As discussed in Section III.A, a “good” seal refers to the proper donning of the headphones with more low-frequency content and corresponds to measurements with greater OASPL. On the other hand, “poor” seals have low-frequency leakage resulting in lower OASPL. Measurements across headphone configurations, represented by the blue circles and red triangles, are grouped by inspection as belonging to a “good” or “poor” seal based on higher or lower OASPL values in Figure 13. Each measurement is within 2 dB of the mean for the assigned group. There are “good” and “poor” seals observed in each of the six headphone configurations for both ears across all measurements. Solid lines in Figure 13 denote the mean over all “good” or “poor” seal measurements over all configurations. Dashed lines bound the shaded region indicating

$\pm 1$  standard deviation across the “good” or “poor” seal measurements. Mean and standard deviation values are noted in Table 3. Visual inspection of Figure 13 confirms the distinct unweighted OASPL trends of the “good” and “poor” seals, with almost a 5-dB difference between a “good” and a “poor” seal.

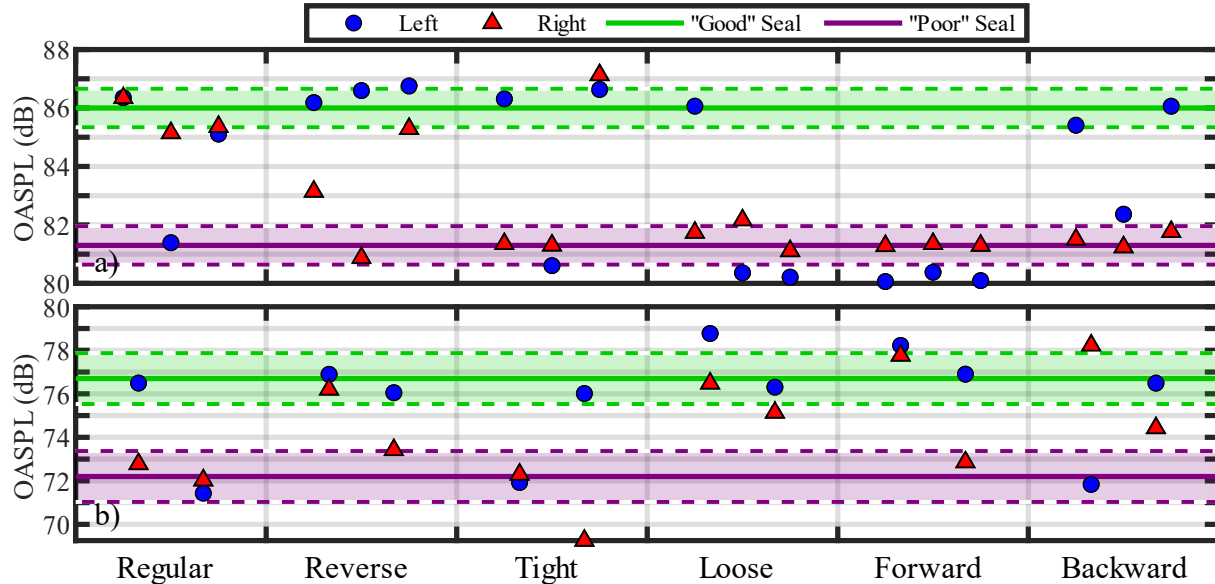


Figure 13. Unweighted OASPL measurements for six headphone configurations (see Figure 3) using setups a) with the DragonFly DAC and b) with the onboard sound card.

Table 3. Unweighted OASPL mean and standard deviation across the “good” and “poor” seal measurements with the DragonFly DAC and with the onboard sound card, as shown in Figure 13.

Setup	Mean “Good” Seal (dB)	St. Dev. “Good” Seal (dB)	Mean “Poor” Seal (dB)	St. Dev. “Poor” Seal (dB)
DragonFly DAC	86.0	0.7	81.3	0.7
Onboard Sound Card	76.7	1.2	72.2	1.4

Similar to the OASPL analysis in Section III.B, A-weighting is applied to the measurements from Figure 13 and shown in Figure 14. The A-weighting decreases the low-frequency content, which in turn decreases the OASPL values. The levels are now more consistent for each channel across all headphone configurations. Mean and standard deviation values in Table 4 are therefore given for each channel as there is no apparent A-weighted OASPL difference between “good” and “poor” seals. Variability for a given channel is small and there is a 1.4-dB offset between channels. The more precise values for A-weighted OASPL suggest that A-weighted metrics are less susceptible to variation due to headphone donning.

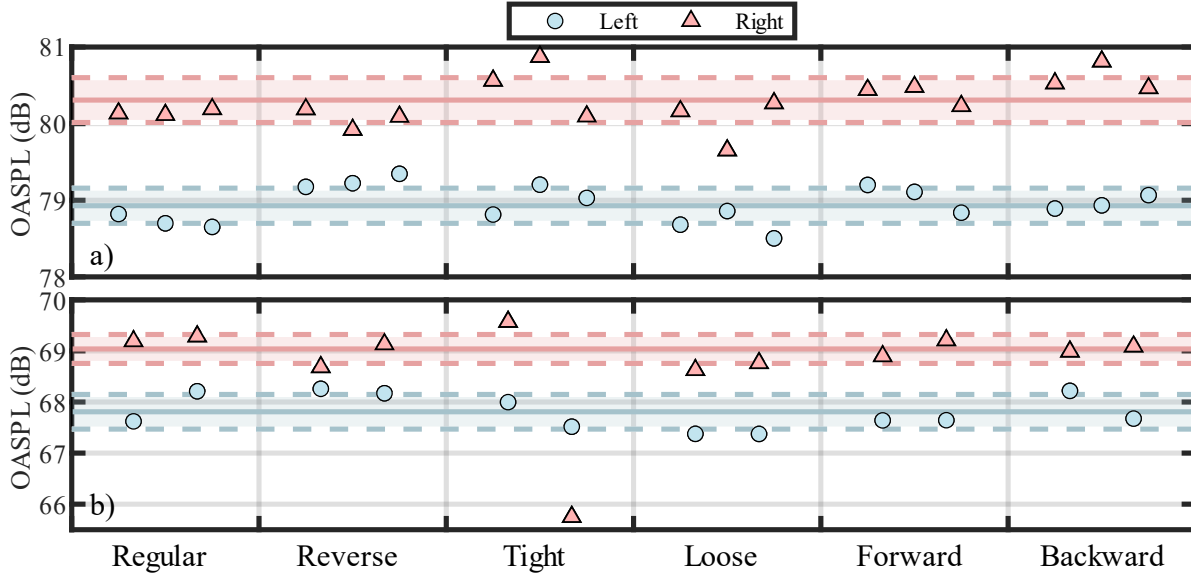


Figure 14. A-weighted OASPL measurements for six headphone configurations (see Figure 3) using setups a) with the DragonFly DAC and b) with the onboard sound card.

Table 4. A-weighted OASPL mean and standard deviation values across measurements of the left and right channels with the DragonFly DAC and with the onboard sound card, as shown in Figure 14. (The second “tight” right channel measurement is excluded as an outlier.)

Setup	Mean Left Channel (dB)	St. Dev. Left Channel (dB)	Mean Right Channel (dB)	St. Dev. Right Channel (dB)
DragonFly DAC	78.9	0.2	80.3	0.3
Onboard Sound Card	67.8	0.3	69.0	0.3

## 2. Calibrated Volume Settings

Results are given here for measurements of 43 stimuli from the VANGARD kiosk test at calibrated volume settings. Histograms in Figure 15 and 16 depict the level of the stimuli relative to the intended level, expressed as a gain in dB, for ASEL and  $LA_{\max}$ . Figure 15 shows results with the computer set to a volume of 21 with the DragonFly DAC and Figure 16 at a volume of 47 with the onboard sound card. The distributions for the left and right channels are similar, with peaks for the right channels occurring at a greater gain value than the left. This trend is similar to what is observed in Figure 14 and is likely due to the measurement system rather than the kiosk. In any case, the majority of tested stimuli are within 2 dB of the intended ASEL and  $LA_{\max}$ . For ASEL, 88% of tested stimuli are within  $\pm 2$  dB with the DragonFly DAC at a volume of 21 and 83% with the onboard sound card at a volume of 47. For  $LA_{\max}$ , 93% of the tested stimuli are within  $\pm 2$  dB both with the DragonFly DAC at a volume of 21 and with the onboard sound card at a volume of 47. These results support those given in Section III.B and provide confidence bounds for the calibrated volume setting.

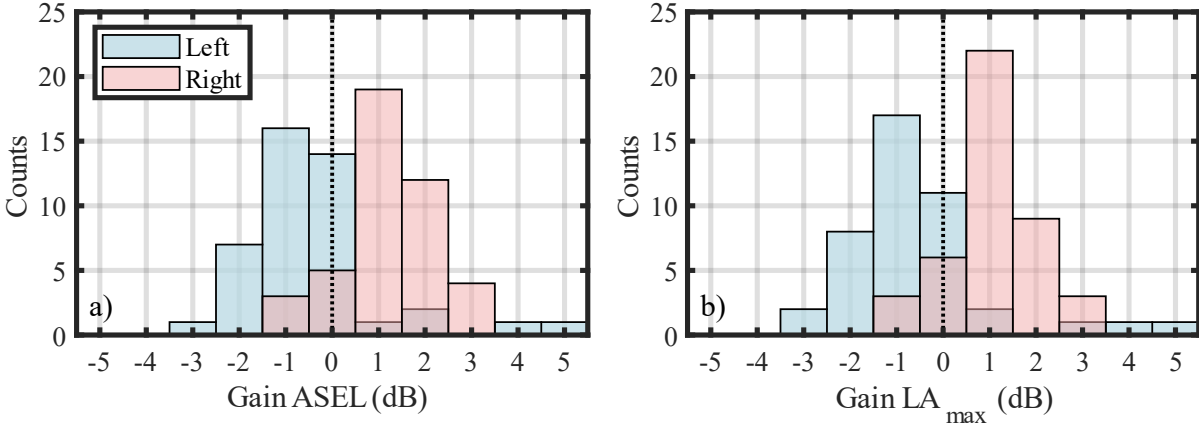


Figure 15. Histograms of the gain for a) ASEL and b) LA<sub>max</sub> measurements relative to the intended level for left and right channels with the DragonFly DAC and the volume set to 21.

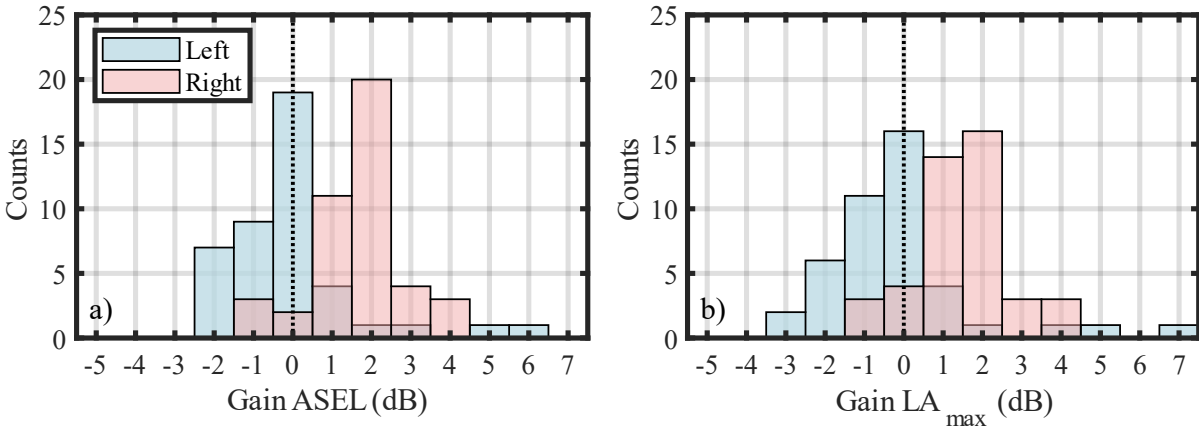


Figure 16. Histograms of the gain for a) ASEL and b) LA<sub>max</sub> measurements relative to the intended level for left and right channels with the onboard sound card and the volume set to 47.

#### IV. CONCLUDING REMARKS

This document demonstrates a method to calibrate kiosks for psychoacoustic testing. The calibrated kiosks consist of a Microsoft Surface Go 3 and a pair of Sennheiser HD 280 headphones. Calibration measurements were made using a manikin head with in-ear microphones. Results from a pink noise signal revealed a consistent spectral shape across volume settings and a reasonably flat spectral response, particularly from 200 to 2,000 Hz. Measurements made with VANGARD test stimuli determined a volume setting of 21 with the DragonFly DAC and 47 with the onboard sound card to achieve intended playback ASEL and LA<sub>max</sub> within  $\pm 2$  dB for 83% to 93% of stimuli tested. While a similar volume setting is likely for other psychoacoustic tests with the kiosks, verifying the levels using elements of this calibration process is recommended.

## References

- Bellows, S. D. (2023). *Acoustic Directivity: Advances in Acoustic Center Localization, Measurement Optimization, Directional Modeling, and Sound Power Spectral Estimation* [PhD Dissertation, Brigham Young University]. [Theses and Dissertations](#), 10021.
- DIY-Audio-Heaven (2023). *HD250 Linear II*. (published July 12, 2017, updated September 22, 2023). <https://diyaudioheaven.wordpress.com/headphones/measurements/brands-s-se/hd250-linear-ii/> (Last accessed 08/21/2025)
- Gee, K. L., Russavage, P. B., Neilsen, T. B., Swift, S. H., and Vaughn, A. B. (2018). “Subjective rating of the jet noise crackle percept,” *J. Acoust. Soc. Am.* 144, EL40–EL34.
- Hirvonen, T., Vaalgamaa, M., Backman, J., and Karjalainen, M. (2003). “Listening test methodology for headphone evaluation,” *AES Convention 114*, 5736.
- Johansson, O., Lindfors, D., Schönfeld, S., Zalyaletdinov, P., and Lukic B. (2012). “A sound sketch approach for auralization of the interior sound of a high speed train,” *ÅForsk Ref. nr.* 09-236.
- Kinnunen, T. (2022). *Headphone Development Research* [Master’s Thesis, Tampere University of Applied Sciences]. Theseus. [https://www.theseus.fi/bitstream/handle/10024/749081/Kinnunen\\_Teemu.pdf](https://www.theseus.fi/bitstream/handle/10024/749081/Kinnunen_Teemu.pdf) (Last accessed 08/21/2025)
- Krishnamurthy, S., Rizzi, S., Biziorek, R., Czech, J., Berg, J., Tannler, D., Bean, D., Ayrapetyan, A., Nguyen, A., and Wivagg, J. (2024). “Remotely Administered Psychoacoustic Test for sUAS Noise to Gauge Feasibility of Remote UAM Noise Study,” *SAE Int. J. Adv. Curr. Pract. Mobil.* 6(2), 986–999.
- NASA (2024). “Remotely Administered Psychoacoustic Test for Advanced Air Mobility Noise Human Response,” *Federal Register*, 2024-05569, (89 FR 18973), [Notice: 24-018].
- Ooi, K., Xie, Y., Lam, B., and Gan, W. S. (2021). “Automation of binaural headphone audio calibration on an artificial head,” *MethodsX* 8, 101288.
- Pickens, A. W., Robertson, L. D., Smith, M. L., Zheng, Q., and Song, S. (2018). “Headphone evaluation for app-based automated mobile hearing screening,” *Int. Arch. Otorhinolaryngol.* 22(04), 358–363.
- Poldy, J. C. (2001). Headphones. In Borwick, J. (ed.) *Loudspeaker and Headphone Handbook* (3rd ed.) (pp. 585–686), CRC Press, London.
- Rizzi, S. A., Huff, D. L., Boyd, D. D., Bent, P., Henderson, B. S., Pascioni, K. A., Sargent, D. C., Josephson, D. L., Marsan, M., He, H., Snider, R. (2020). “Urban Air Mobility Noise: Current Practice, Gaps, and Recommendations,” *NASA/TP-20205007433*, NASA Langley Research Center, Hampton, VA.
- Schultz, T. J. (1978). “Synthesis of social surveys on noise annoyance,” *J. Acoust. Soc. Am.* 64(2), 377–405.
- Sennheiser electronic SE & Co. KG. (2025). *HD 280 PRO*. <https://www.sennheiser.com/en-us/catalog/products/headphones/hd-280-pro/hd-280-pro-506845> (Last accessed 08/21/2025)

- Sørensen, C. B., Adams, T. B., Pedersen, E. R., Nielsen, J., and Schmidt, J. H. (2023). “AMTAS™ and user-operated smartphone research application audiometry—An evaluation study,” *PLoS ONE* 18(9), e0291412.
- Strong, W. J. and Plitnik, G. R. (2007). *Sensory Perceptions and Their Physical Bases*, in *Music Speech Audio* (3rd ed.) (pp. 111–112), Brigham Young University Publications, Provo.
- Torija Martinez, A. J. and Li, Z. (2020). “Metrics for assessing the perception of drone noise,” in *Forum Acusticum*, 3163–3168.
- Torres-Russotto, D., Landau, W.M., Harding, G.W., Bohne, B.A., Sun, K., and Sinatra, P. M. (2009). “Calibrated Finger Rub Auditory Screening Test (CALFRASST),” *Neurology* 72(18), 1595–1600.
- Whiting, J. K. (2018). *Development of a Real-Time Auralization System for Assessment of Vocal Effort in Virtual-Acoustic Environments* [Master’s Thesis, Brigham Young University]. *Theses and Dissertations*, 7056.
- World Health Organization (2018). “Environmental noise guidelines for the European Region,” World Health Organization, Regional Office for Europe, Copenhagen, Denmark. <https://iris.who.int/handle/10665/279952> (Last accessed 08/21/2025)
- Xie, Z., Shen, J., Liu, Y., and Rao, D. (2009). “Calibration of headphones and earphone with KEMAR,” *2nd International Congress on Image and Signal Processing*, IEEE, pp. 1–4.

## Acknowledgments

Work conducted in support of the NASA Revolutionary Vertical Lift Technology Project. The author expresses appreciation to D. Begault for his help with calibration efforts and the following reviewers of this manuscript: A. Allen, S. Krishnamurthy, and R. Cabell.

## Appendix A: VANGARD Kiosk Test Calibration Stimuli

This appendix contains additional plots from calibration testing with the VANGARD kiosk test stimuli. These figures are included for completeness and as additional examples of the discussion in Section III.C. Figure 17 to 18 show the A-weighted SPL versus time, similar to Figure 8, of Signals 2 through 3 from Table 1. Figure 20 and 20 show the gain of  $LA_{max}$  versus volume setting, similar to the ASEL results in Figure 9 and 10. Notably, Signal 2 diverges at the lowest volume settings in Figure 21 because it is the quietest signal and approaches the noise floor of the microphone and possibly the room itself at the start and end of the signal, as shown in Figure 17.

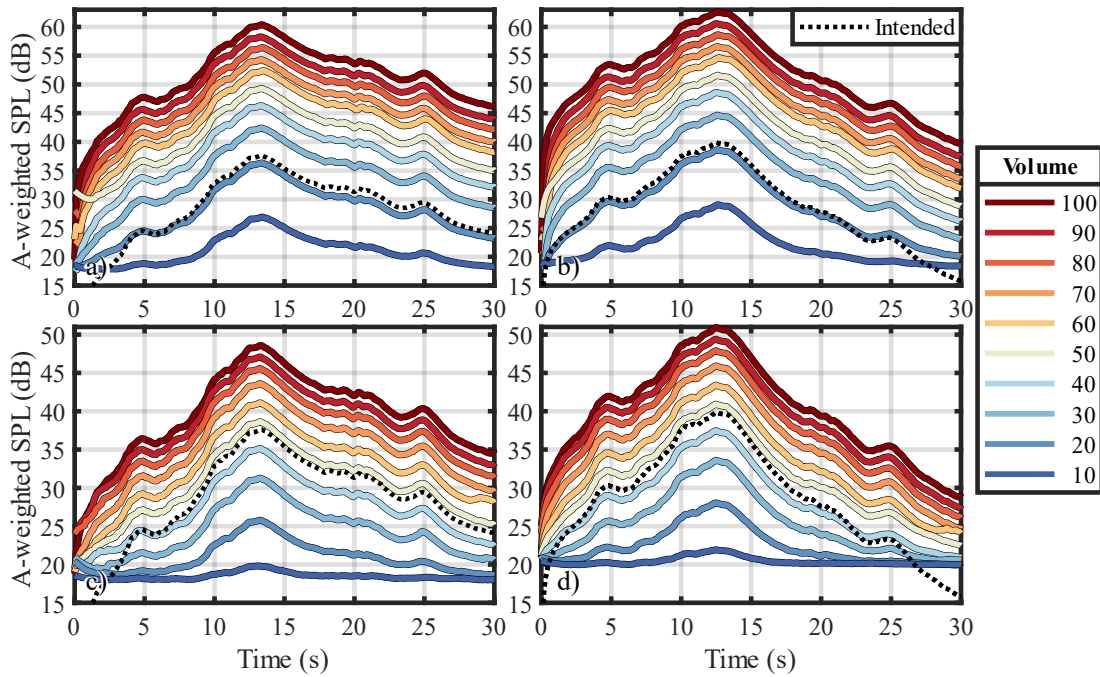


Figure 17. VANGARD test Signal 2 measured versus intended slow, A-weighted SPL versus time for the a) left and b) right channels with the DragonFly DAC and c) left and d) right channels with the onboard sound card.

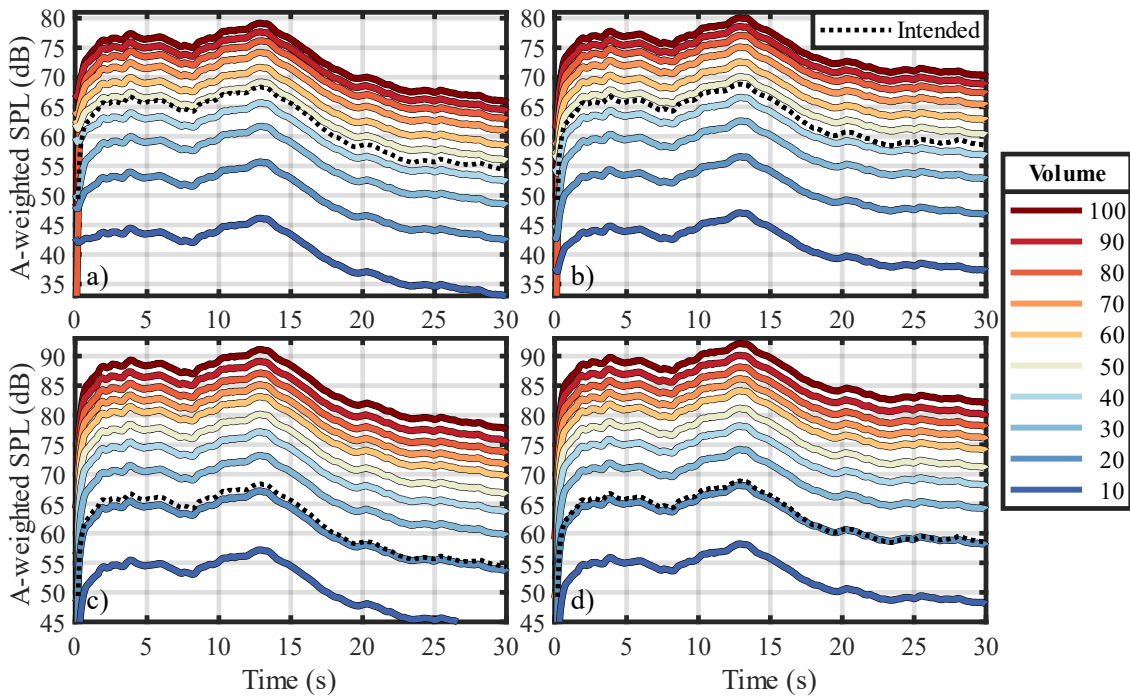


Figure 18. VANGARD test Signal 3 measured versus intended slow, A-weighted SPL versus time for the a) left and b) right channels with the DragonFly DAC and c) left and d) right channels with the onboard sound card.

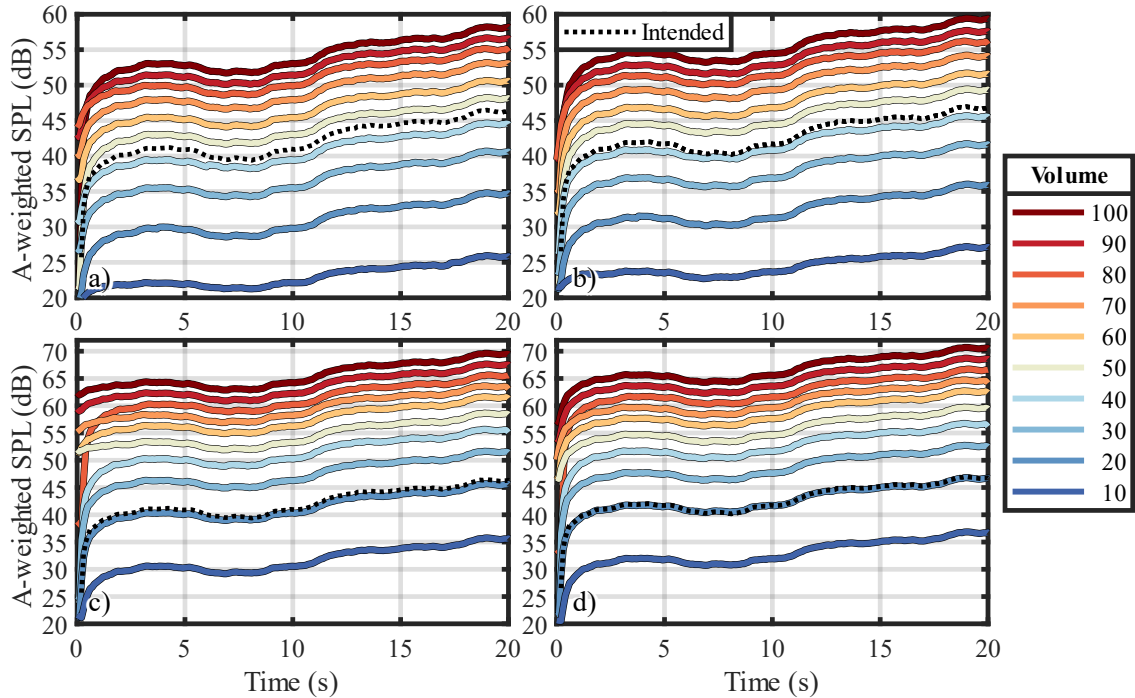


Figure 19. VANGARD test Signal 4 measured versus intended slow, A-weighted SPL versus time for the a) left and b) right channels with the DragonFly DAC and c) left and d) right channels with the onboard sound card.

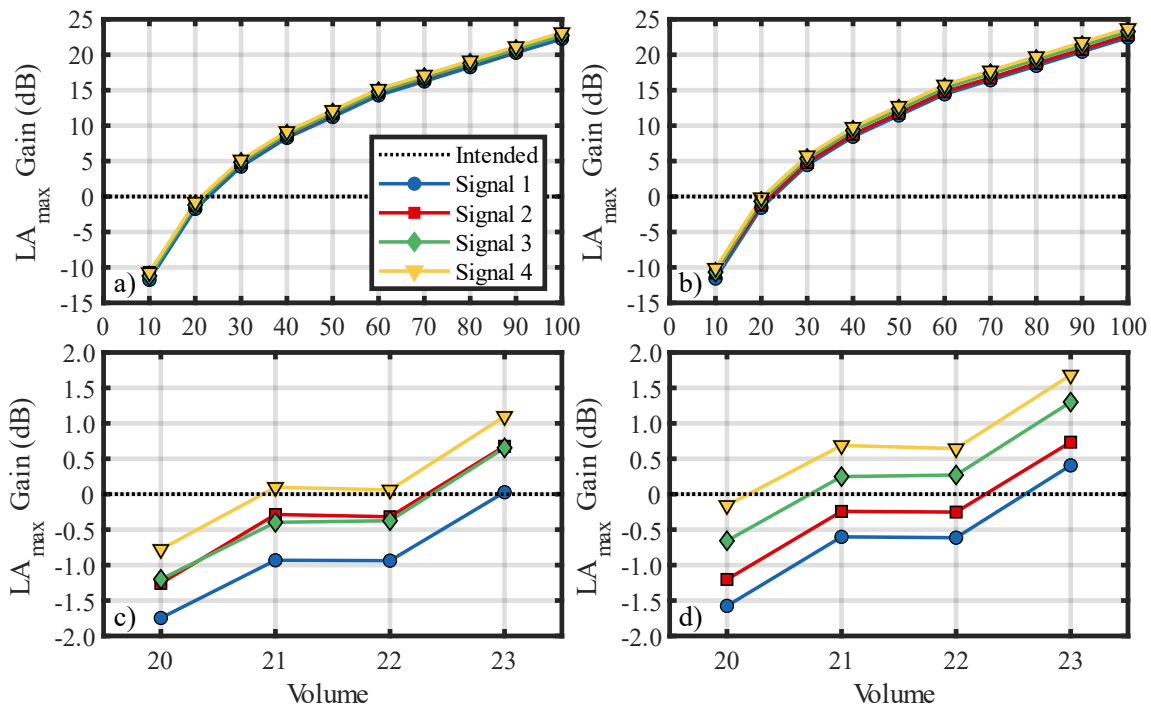


Figure 20. Gain of LA<sub>max</sub> measurements with the DragonFly DAC relative to the intended level in volume increments of 10 for the a) left and b) right channels and in volume increments of 1 near ASEL gain of 0 dB for the c) left and d) right channels.

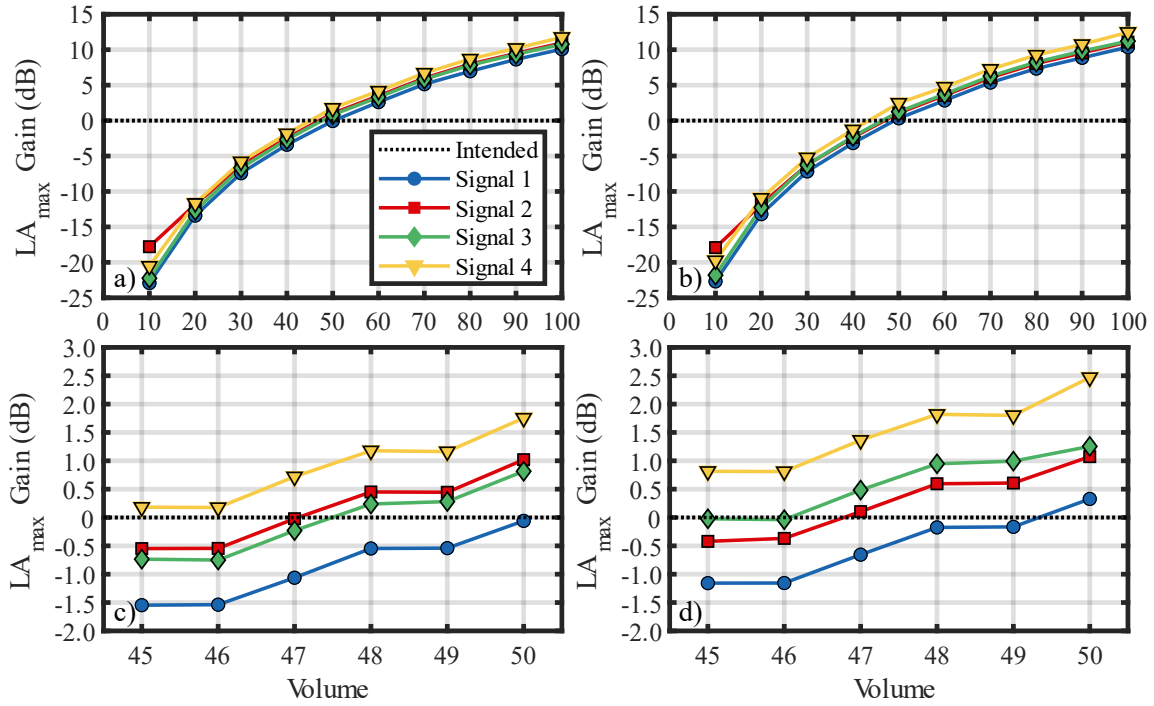


Figure 21. Gain of  $LA_{max}$  measurements with the onboard sound card relative to the intended level in volume increments of 10 for the a) left and b) right channels and in volume increments of 1 near ASEL gain of 0 dB for the c) left and d) right channels.