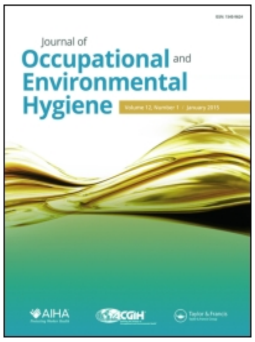

Taylor & Francis
Taylor & Francis Group

Journal of Occupational and Environmental Hygiene

ISSN: 1545-9624 (Print) 1545-9632 (Online) Journal homepage: http://www.tandfonline.com/loi/uoeh20

Inter-laboratory comparison of three earplug fit-test systems

David C. Byrne, William J. Murphy, Edward F. Krieg, Robert M. Ghent, Kevin L. Michael, Earl W. Stefanson & William A. Ahroon

To cite this article: David C. Byrne, William J. Murphy, Edward F. Krieg, Robert M. Ghent, Kevin L. Michael, Earl W. Stefanson & William A. Ahroon (2017) Inter-laboratory comparison of three earplug fit-test systems, Journal of Occupational and Environmental Hygiene, 14:4, 294-305, DOI: 10.1080/15459624.2016.1250002

To link to this article: http://dx.doi.org/10.1080/15459624.2016.1250002

Accepted author version posted online: 27 Oct 2016.
Published online: 27 Oct 2016.

JOURNAL OF OCCUPATIONAL AND ENVIRONMENTAL HYGIENE
2017, VOL. 14, NO. 4, 294–305
http://dx.doi.org/10.1080/15459624.2016.1250002

Inter-laboratory comparison of three earplug fit-test systems

David C. Byrne[a], William J. Murphy[b], Edward F. Krieg[b], Robert M. Ghent[c], Kevin L. Michael[d], Earl W. Stefanson[e], and William A. Ahroon[e]

[a]NIOSH - Robert A. Taft Laboratories, Cincinnati, Ohio; [b]NIOSH, Cincinnati, Ohio; [c]Honeywell Safety Products, San Diego, California; [d]Michael & Associates, State College, Pennsylvania; [e]U.S. Army Aeromedical Research Laboratory, Fort Rucker, Alabama

ABSTRACT

The National Institute for Occupational Safety and Health (NIOSH) sponsored tests of three earplug fit-test systems (NIOSH HPD Well-Fit, Michael & Associates FitCheck, and Honeywell Safety Products VeriPRO). Each system was compared to laboratory-based real-ear attenuation at threshold (REAT) measurements in a sound field according to ANSI/ASA S12.6-2008 at the NIOSH, Honeywell Safety Products, and Michael & Associates testing laboratories. An identical study was conducted independently at the U.S. Army Aeromedical Research Laboratory (USAARL), which provided their data for inclusion in this article. The Howard Leight Airsoft premolded earplug was tested with twenty subjects at each of the four participating laboratories. The occluded fit of the earplug was maintained during testing with a soundfield-based laboratory REAT system as well as all three headphone-based fit-test systems. The Michael & Associates lab had the highest average A-weighted attenuations and smallest standard deviations. The NIOSH lab had the lowest average attenuations and the largest standard deviations. Differences in octave-band attenuations between each fit-test system and the American National Standards Institute (ANSI) sound field method were calculated ($\text{Atten}_{\text{fit-test}}$ - $\text{Atten}_{\text{ANSI}}$). A-weighted attenuations measured with FitCheck and HPD Well-Fit systems demonstrated approximately ±2 dB agreement with the ANSI sound field method, but A-weighted attenuations measured with the VeriPRO system underestimated the ANSI laboratory attenuations. For each of the fit-test systems, the average A-weighted attenuation across the four laboratories was not significantly greater than the average of the ANSI sound field method. Standard deviations for residual attenuation differences were about ±2 dB for FitCheck and HPD Well-Fit compared to ±4 dB for VeriPRO. Individual labs exhibited a range of agreement from less than a dB to as much as 9.4 dB difference with ANSI and REAT estimates. Factors such as the experience of study participants and test administrators, and the fit-test psychometric tasks are suggested as possible contributors to the observed results.

KEYWORDS
Fit-test; hearing protection; noise reduction rating

Introduction

Hearing protector attenuation is determined either by conducting listening tests with human subjects or by direct acoustical measurement.[1] The U.S. Environmental Protection Agency (EPA) requires the single-number noise reduction rating (NRR) to be shown on the label of each hearing protector sold in the United States.[2] The NRR is based upon human subject testing conducted in the laboratory with a panel of listeners. Hearing tests are administered to these individuals with and without the protectors in place. This test method is referred to as real-ear attenuation at threshold (REAT); the decibel difference between the open-ear threshold and the occluded-ear threshold indicates the amount of sound attenuation provided.

Unfortunately, testing hearing protection devices (HPDs) for rating purposes bears little resemblance to their use in the real world. Attenuation values obtained from laboratory NRR tests tend overestimate the average amount of protection received by individuals during everyday use.[3] Well-trained and highly motivated individuals are normally used as subjects in the laboratory, whereas the typical noise-exposed worker may not receive the necessary training or assistance with fitting the protectors. Also, workers may intentionally compromise the fit of their protectors to provide what they deem

CONTACT David C. Byrne dbyrne@cdc.gov NIOSH - Robert A. Taft Laboratories, 1090 Tusculum Avenue, Mail Stop C-27, Cincinnati, OH 45226.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uoeh.

is sufficient protection while maximizing their ability to communicate. While the NRR may provide a general estimate of how much noise reduction a particular hearing protector is capable of providing, many purchasers of hearing protection mistakenly assume that a higher NRR is always better. In reality, the labeled NRR is intended to inform the consumer about the potential performance of a protector and is not an indicator of the performance for any individual. A significant shortfall of the NRR is that the attenuation for an individual user cannot be predicted from the sample statistic derived from laboratory data.

The American National Standards Institute (ANSI) Subcommittee for Noise, Working Group 11 (S12/WG11) conducted a six-laboratory comparison study of the Method A "experimenter-supervised fit" and Method B "subject fit" protocols for REAT described in the ANSI/ASA S12.6-1997 (R2002) standard.[4,5] The Method A attenuations were greater than the Method B attenuations and the inter-laboratory variance was greater for Method A. The many factors affecting hearing protector attenuation values motivated the development of individual fit-testing methods.

Fit-testing determines how well an individual fits a specific hearing protector. Fit-testing is recommended as a "best practice" for hearing loss prevention programs.[6–8] In 1976, the National Institute for Occupational Safety and Health (NIOSH) published a report about the fit-testing system developed by The Pennsylvania State University.[9] The Penn State system used modified large-volume earmuffs with transducers mounted in the ear cups to present the stimuli. Several racks of equipment installed in a mobile testing van were required to conduct the fit-test. Laptop computers and high-performance digital audio have replaced the equipment racks and provided much-needed portability to take the fit-test to the worker.

Several fit-test systems are now commercially available. REAT-under-headphone fit-test systems replicate the ANSI laboratory test by presenting the stimuli via headphones rather than loudspeakers in a diffuse sound field. Occluded and open-ear hearing thresholds are measured at specific test frequencies. Hearing protector attenuation is estimated from differences between the occluded and open-ear hearing thresholds at each test frequency. The alternating binaural loudness balance (ABLB) fit-test system was developed to overcome the effects of ambient noise that might mask the open-ear threshold. In this case, the user sequentially performs a loudness balance between open-ear and occluded conditions to estimate the attenuation of the protector. The microphone-in-real-ear (MIRE) technique may be conducted with probe measurement systems similar to those used for evaluating hearing aid performance. A probe tube is situated between the earplug and the canal wall or may be inserted through a hole drilled through an earplug specifically modified for this purpose. The sound level in the ear canal under the hearing protector is simultaneously measured along with the level outside the earplug. The difference in sound pressure level between the two microphones adjusted for compensation factors between the MIRE and REAT measurement allow for the estimation of the attenuation provided by the earplug. No MIRE-based systems were included in the current study.

This inter-laboratory study compares three hearing protector fit-test systems to the ANSI/ASA S12.6-2008 laboratory standard method for measuring the noise-reducing capabilities of hearing protection devices.[10] NIOSH Taft Laboratories (Cincinnati, OH) contracted with the Howard Leight Acoustical Testing Laboratory at Honeywell Safety Products (San Diego, CA) and the Michael & Associates Psychoacoustics Laboratory (State College, PA) to perform these tests. An identical study was conducted independently at the U.S. Army Aeromedical Research Laboratory (USAARL, Fort Rucker, AL), which provided their data for inclusion in this report. The three fit-test systems that were evaluated included the VeriPRO (Honeywell), FitCheck (Michael & Associates), and HPD Well-Fit (NIOSH). The purpose was to gather comparative data on attenuation measurements obtained with the different systems in order to determine the correlation of fit-testing results with REAT testing in the laboratory and the relative performance across the three field measurement systems.

Methods

Subjects

Each laboratory recruited 20 subjects from the local community to participate in the study. The project was approved by the NIOSH Institutional Review Board (Protocol 10-DART-06XP), the U.S. Army Medical Research and Materiel Command Institutional Review Board (M-10119), and the USAARL Institutional Review Board (2010-050). Subjects signed an informed consent form before testing began.

Per the ANSI S12.6 standard, subjects were required to meet certain inclusion criteria, one of which was normal hearing thresholds (i.e., ≤ 25 dB HL) from 125–8000 Hz, as determined with standard clinical audiometry. Except for volunteers at USAARL, subjects were paid a small stipend for their participation, which typically required two visits to the laboratory consisting of about two hours of testing per visit. Subjects were qualified during the first visit and trained in the psychophysical test method used with the laboratory REAT system. The ANSI standard requires five open-ear tests prior to testing of hearing protection.[10] Each subject's last three practice tests must exhibit open-ear thresholds with a range less than 6 dB at each test frequency. Subjects that did not achieve that level of consistency after re-training were dismissed from the study.

The level of experience with hearing protector use was not considered a criterion for subjects to be included in the study. Because hearing protector fit-testing—like respirator fit-testing—will be performed in the workplace with both highly experienced and inexperienced personnel, subjects were selected without controlling for their prior experiences. At the NIOSH laboratory, four of the subjects were experienced listeners (i.e., audiologists or hearing scientists) and the remaining subjects were recruited from the local community. At the Honeywell laboratory, approximately half of the subjects were recruited from those who had participated in prior studies with the remainder being unfamiliar with hearing protector testing. At the Michael & Associates laboratory, all of the subjects were participants in hearing protector testing that is regularly scheduled as a part of that laboratory's activities. Six subjects at the USAARL facility were experienced listeners (i.e., an audiologist, engineers, and hearing technicians).

Laboratory facilities

Each of the four laboratories has extensive experience testing hearing protection according to the ANSI/ASA S12.6-2008 standard.[10] The standard sets requirements for sound field uniformity, directionality, reverberation time, and ambient noise levels inside the test room. Calibrated one-third octave-band noise stimuli are created by a computer-controlled noise source and passed through an attenuator to a power amplifier and loudspeakers. The subject's head is positioned at the center of the diffuse sound field. Subjects use a hand switch to indicate detection of a stimulus. Thresholds are determined according to an automatic modified Békésy tracking paradigm and are calculated by averaging the midpoints for six ascending and descending traces. Attenuation values are obtained by subtracting the open-ear thresholds from the occluded thresholds at each test frequency.

Field measurement systems

The VeriPRO system uses an ABLB paradigm where the subject first matches the loudness of pure tones at five frequencies (250, 500, 1000, 2000, and 4000 Hz) presented via manufacturer-supplied headphones when no hearing protection is worn.[11] Starting with a set reference tone in the right ear, the tones alternate between ears and the subject adjusts the level in the left ear until it is perceived to be of equal loudness as the tone in the right ear. The level of the reference tone was set to 60 dB SPL in this study. Next, the subject inserts an earplug into the right ear. The presentation level in the right ear is increased (up to 90-dB in this study) and the loudness matching is repeated by adjusting the level of the tone in the left ear. Finally, the subject inserts an earplug into the left ear and performs a loudness balance with both ears occluded and the right ear reference level is fixed (90 dB SPL in this study). The differences between the open-ear and occluded conditions are used to estimate the attenuation at each frequency, and a personal attenuation rating (PAR) is computed for the left and right ears separately. VeriPRO uses attenuations at five frequencies to compute the PAR. The order of frequencies tested was 500, 250, 1000, 2000 and 4000 Hz. The VeriPRO system performs the reliability check (6-dB tolerance) for only the first open-ear frequency (500 Hz) and does not test reliability at other open-ear frequencies.

The FitCheck system determines open-ear and occluded thresholds for one-third octave-bands of noise using the modified Békésy threshold-tracking paradigm. FitCheck uses large volume headphones to present stimuli instead of loudspeakers as used in the laboratory REAT system. Like the laboratory system, subjects increase and decrease the stimulus presentation level with a response switch. Threshold is calculated after a predetermined number of up and down cycles and testing progresses to the next frequency. FitCheck measures attenuations at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. Although individual testing of each ear is possible with this system, the current study used a binaural (diotic) signal presentation, which produced a single attenuation value at each frequency. At each frequency, FitCheck corrects the headphone output level to the laboratory diffuse sound field levels based upon a series of linear regressions determined at the Michael & Associates laboratory. The corrected FitCheck headphone presentation levels measured with the FitCheck were used to estimate PAR.

The HPD Well-Fit system is a REAT-under-headphone fit-test system. A principal difference between HPD Well-Fit and the ANSI sound-field method is that HPD Well-Fit uses a Method of Adjustment paradigm. The subject controls the presentation level of a dichotic, pulsed one-third octave-band of noise with the scroll wheel of a computer mouse. Once the subject determines the level at which the noise is barely audible, he/she clicks one of the mouse buttons. The software then randomly increases the level of the stimulus between 10 and 20 dB and the subject repeats the adjustment procedure until three threshold identifications agree to within a tolerance range of 6 dB.

Table 1. Manufacturer's reported labeled attenuation as a function of frequency for the Howard Leight Airsoft earplug. The octave-band real-ear attenuation at threshold (REAT) data were obtained by the Howard Leight Acoustical Testing Laboratory and result in an NRR of 27 dB.

Frequency (Hz)	125 Hz	250 Hz	250 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	NRR
Mean Attenuation (dB)	22.9	24.8	28.9	31.8	35.4	43.1	38.1	27
Standard Deviation (dB)	1.6	1.7	2.0	2.0	2.7	2.9	3.0	

Thresholds were obtained at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz for both the open-ear and occluded conditions. Differences between the occluded and open-ear thresholds are used to estimate the attenuation. The HPD Well-Fit system may be used with any circumaural headphone with sufficient ear-cup volume. For this study, the FitCheck headphones supplied with the FitCheck system were used at all four participating laboratories. Output levels were calibrated for HPD Well-Fit and attenuations were not corrected as was done with FitCheck.

Standard earplug

The Howard Leight Airsoft premolded earplug used for this study has a Noise Reduction Rating (NRR) of 27 dB (see Table 1 for octave band attenuations). The Airsoft earplug has a stem and four flanges that surround an air-filled bladder. The Airsoft earplug can be easily fit by the subjects and was expected to require a minimal amount of training for proper insertion.

Test procedure

After conducting the initial subject qualification and REAT system training per ANSI S12.6, testing was performed with each of the four methods. Training on each of the three fit-test systems was provided according to the manufacturers' instructions, and the standard laboratory instructions were given for the ANSI S12.6 tests. The occluded and open-ear tests were conducted in accordance with Section 7.3 of the ANSI S12.6 standard, whereby one-half of the subjects received an open-ear test first and the other half received an occluded test first.[10] Because a right/left ear test order is hard-coded into the VeriPRO loudness balance test paradigm, it was always administered immediately prior to the occluded test runs. Otherwise, the order of the attenuation tests was randomized for the other two fit-test systems and the laboratory method. One example order with open-ear tests being first is open-ear ANSI, WellFit, FitCheck VeriPRO then occluded VeriPRO FitCheck, WellFit, and ANSI. Another example test order with occluded tests first would be open-ear VeriPRO, occluded VeriPRO FitCheck Well-Fit and ANSI then open-ear ANSI WellFit and FitCheck. Once a subject inserted both earplugs, they were not permitted to adjust the plugs until all of the occluded tests were completed. Earplugs were removed and reinserted for the second trial and testing proceeded with a different sequence of systems.

As described above, all four laboratories had double-walled sound-isolating booths for conducting the ANSI laboratory tests. At the NIOSH laboratory, the VeriPRO system was tested in a quiet office space and the FitCheck and HPD Well-Fit systems were tested in a small single-walled audiometric booth [approximately 1.2 m (4′) L × 1.0 m (3.4′) W × 2.0 m (6.5′) H]. At the Honeywell lab, all fit-testing was conducted in the single-wall control room adjacent to the ANSI test room. Michael & Associates conducted their fit-testing sessions in a quiet office space. At USAARL, all tests were performed in the double-walled booth used for ANSI REAT testing.

Statistical methods

A mixed effects linear model was used to analyze the data. The SAS PROC MIXED (SAS Release 9.3, SAS Institute, Inc., Cary, NC) procedure was used to estimate the model. The method of estimation was residual maximum likelihood. The fixed effects in the model were laboratory, method, and the laboratory by method interaction. The random effects were subject within laboratory and subject within laboratory by method. Separate variances were estimated for each laboratory. The residual variance was estimated separately for each combination of laboratory and method.

Results

The A-weighted attenuation, Awt_N, was calculated for each subject according to the following equation:

$$Awt_N = 10log_{10}\sum_{f}^{N}10^{L_{Af}/10} - 10log_{10}\sum_{f}^{N}10^{(L_{Af}-A_f)/10}, \quad (1)$$

where L_{Af} is 100 dB plus the A-weighting factor for each stimulus frequency (f) and A_f is the attenuation measure by the fit-testing system at each frequency. For the FitCheck, HPD Well-Fit, and ANSI (laboratory) systems, the frequencies tested and used in the estimate were 125,

Table 2. Mean attenuation and standard deviation by frequency and lab for the ANSI test. The A-weighted attenuation (Awt_7) was calculated for each subject's trial and the mean is across all subjects. The standard deviations are determined after averaging each subject's trials first.

Lab	Quantity	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	A-weighted Attenuation	NRR
Honeywell	Mean Attenuation (dB)	21.5	21.0	21.8	22.2	27.3	28.0	33.2	23.4	4.4
	Std Deviation (dB)	9.4	9.8	10.8	8.9	7.9	7.3	10.2	7.9	
Michael & Associates	Mean Attenuation (dB)	28.7	25.2	29.1	27.8	34.0	34.3	42.1	30.1	19.5
	Std Deviation (dB)	4.2	4.6	4.9	4.9	3.9	5.7	4.2	3.6	
NIOSH	Mean Attenuation (dB)	15.8	16.8	18.3	19.7	24.6	24.4	30.2	20.5	− 1.1
	Std Deviation (dB)	10.7	9.9	11.5	10.2	7.9	10.7	12.6	9.5	
USAARL	Mean Attenuation (dB)	23.5	22.5	23.4	21.0	30.0	28.1	34.0	23.5	4.8
	Std Deviation (dB)	10.3	10.4	11.0	8.8	6.6	7.9	8.3	7.7	

250, 500, 1000, 2000, 4000, and 8000 Hz. For the VeriPRO system, the frequencies used were 250, 500, 1000, 2000, and 4000 Hz. For each system and subject trial, subtracting the ANSI (laboratory) results from the fit-test system results yielded the difference. Trials for each subject were averaged and the results for each lab also were averaged.

In Table 2, the mean attenuations and standard deviations for the ANSI laboratory measurements were calculated both by frequency and the 7-frequency A-weighted attenuation, Awt_7. The Michael & Associates laboratory had the highest average Awt_7 value of 30.1 dB, followed by USAARL and Honeywell with 23.5 and 23.4 dB, respectively. The NIOSH laboratory had a mean Awt_7 of 20.5 dB. The Michael & Associates subjects had the smallest standard deviations and NIOSH subjects had the largest standard deviations. The standard deviations reflect the consistency of the subject panel to achieve the same attenuation. The Michael & Associates subjects regularly participated in hearing protector testing panels. The subjects at the other labs were a mixture of experienced subjects and inexperienced subjects. No criteria were applied with regards to prior experience or ability to fit the earplug.

Table 3 examines the possibility that subjects may have learned to better fit the earplug between trial 1 and trial 2. For each test method, ANSI, FitCheck, Well-fit, and VeriPRO, the differences between the first and second trials were determined and the average and 95% confidence intervals were calculated using the normfit function within the MATLAB statistics toolbox (MATLAB release 2016a, Mathworks Inc., Natick, MA). The confidence interval for the VeriPRO Right ear measurement in the Honeywell laboratory was the only one that was significantly different from zero. The asymmetry of the confidence intervals in some of the cases may indicate that the first or second trial was better or worse–particularly Honeywell VeriPRO Left ear, NIOSH FitCheck and NIOSH Well-fit. If the ANSI data are used as the standard, then none of the labs had subjects that exhibited a definite bias for first trial to second trial. In other words, the subjects did not fit the plug significantly better the second time compared to the first trial. This result would be expected because no training was provided between trials.

In Figure 1, the mean attenuations and standard deviations (N = 20) at each frequency for all tests are shown. Individual ear results are shown for the VeriPRO system because the test paradigm requires the right and left ears always to be tested separately. Some trends in the data are evident. First, in all except the NIOSH laboratories, the HPD Well-Fit (square symbols) exhibited a lower average attenuation level than the other systems at 125 Hz. The VeriPRO data (red and blue triangles) were below the ANSI results at 2000 and 4000 Hz at all four test labs. The VeriPRO attenuations at 500 and 1000 Hz were also less than the ANSI results at the NIOSH lab. In the Michael & Associates lab, the standard deviations tended to be less than those obtained at the other three laboratories. The variances of the attenuations observed in Figure 1 were tested using the Pitman Morgan test for significant differences when compared to the variance measured for the ANSI test.[12] The error bars are marked with an asterisk that indicate whether the variance for a lab, method and frequency was significantly different at the $p < 0.05$ probability level. For the Honeywell lab, none of the variances were significantly different compared to the variance for the ANSI method. For the Michael & Associates lab, the variances were significant larger for the HPD Well-Fit at

Table 3. Mean and 95% confidence interval for the difference between A-weighted attenuation values (first trial minus second trial for each fit-test system and laboratory. The bolded entry indicates that the confidence interval did not include zero.

	ANSI	FitCheck	HPD Well-Fit	VeriPRO Left	VeriPRO Right
Honeywell	1.2 (−2.4, 4.8)	2.6 (−0.1, 5.3)	− 0.3 (−3.2, 2.6)	2.3 (−0.9, 5.5)	**4.8 (1.1, 8.6)**
MA	− 0.0 (−1.8, 1.8)	0.1 (−1.1, 1.4)	0.4 (−2.0, 2.7)	1.8 (−1.0, 4.7)	1.2 (−1.4, 3.7)
NIOSH	0.5 (−3.0, 3.9)	− 3.3 (−8.6, 1.9)	− 1.8 (−5.4, 1.8)	0.6 (−5.0, 6.2)	− 1.4 (−5.1, 2.3)
USAARL	2.0 (−0.1, 4.0)	− 0.1 (−2.7, 2.5)	1.6 (−1.8, 5.1)	− 1.1 (−4.2, 2.0)	1.3 (−2.7, 5.2)

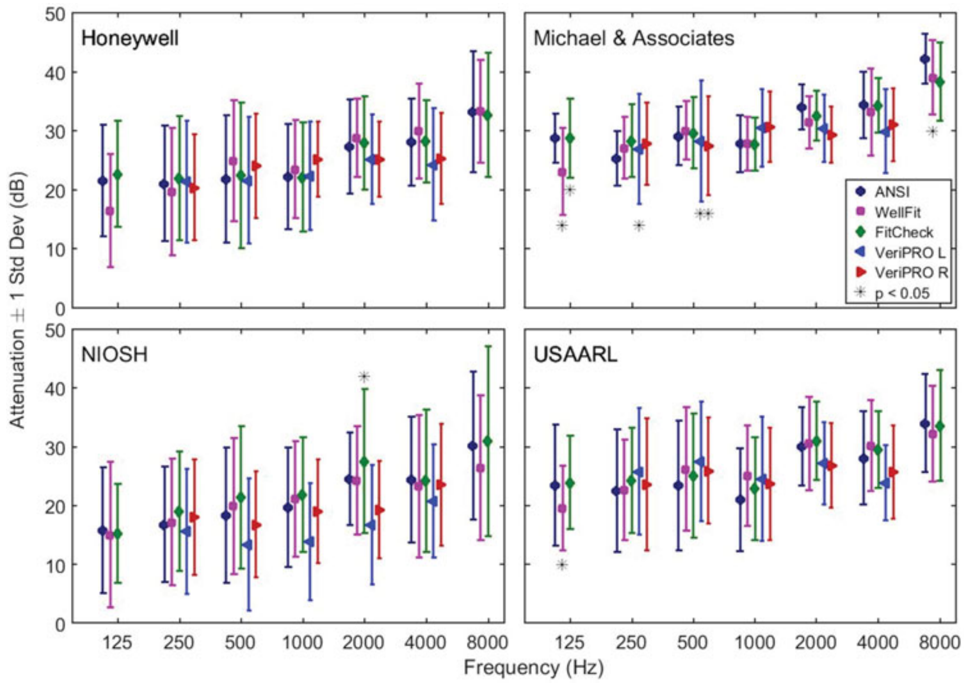

Figure 1. Mean attenuation values by frequency for each laboratory and fit-test system. Error bars indicate ±1 standard deviation. Only three data points are shown at 125 Hz and 8000 Hz because the VeriPRO system does not test these frequencies. Within the labs, frequencies where the variance is significantly different ($p < 0.05$) from the corresponding variance of the ANSI test are marked with an asterisk.

125 and 800 Hz, for FitCheck at 125 Hz, for VeriPRO left ear at 250 and 500 Hz, and for VeriPRO Right at 500 Hz. For the NIOSH lab, only the variance for FitCheck at 2000 Hz was significantly different. For the USAARL lab, only the variance for FitCheck at 125 Hz was significantly different.

Figure 2 shows the averaged A-weighted attenuations determined with the available data for each laboratory and fit-test system. The first set of symbols in each laboratory (circles) match the A-weighted attenuations for the 7-frequencies tested and reported in Table 2. The error bars are the 95% confidence intervals determined from the statistical model. As will be shown in Figure 4, statistically significant differences between the fit-test system and the ANSI method can be identified when the mean value of ANSI and the fit-test system do not fall within the respective confidence intervals (e.g., Honeywell ANSI and VeriPRO Left vs. Honeywell ANSI and VeriPRO Right).

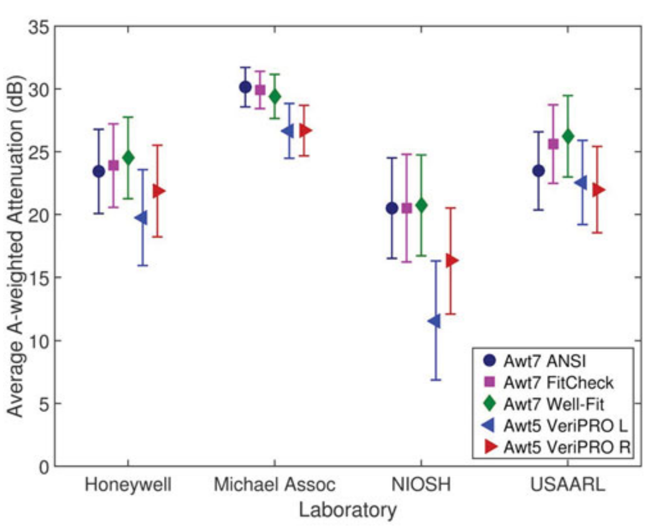

Figure 2. A-weighted attenuations computed for each fit-test system and laboratory. Error bars represent 95% confidence intervals.

Figure 3 shows the correlation between each of the three fit-test systems and the corresponding ANSI laboratory results. Different symbol shapes represent the four participating test laboratories. Overall attenuation values represent the average of seven frequencies (125–8000 Hz) for the FitCheck and HPD Well-Fit systems when compared to the ANSI method, whereas the VeriPRO results are compared using five frequencies (250–4000 Hz). Correlation coefficients for the FitCheck and HPD Well-Fit results were similar ($r = 0.84$ and $r = 0.86$, respectively). Correlation coefficients for the left and right ears of the VeriPRO results were $r = 0.49$ and $r = 0.55$, respectively. The majority of the A-weighted attenuation values for the VeriPRO underestimated the ANSI attenuation results. From the frequency-by-frequency comparisons in Figure 1, the attenuations for 2000 and 4000 Hz tended to be underestimated by the VeriPRO for both the right and left ear. Low attenuation values at low frequencies tend to indicate a poorly fit earplug and the consequential low overall attenuation. For subjects at the Michael & Associates lab, the VeriPRO vs. ANSI distributions are

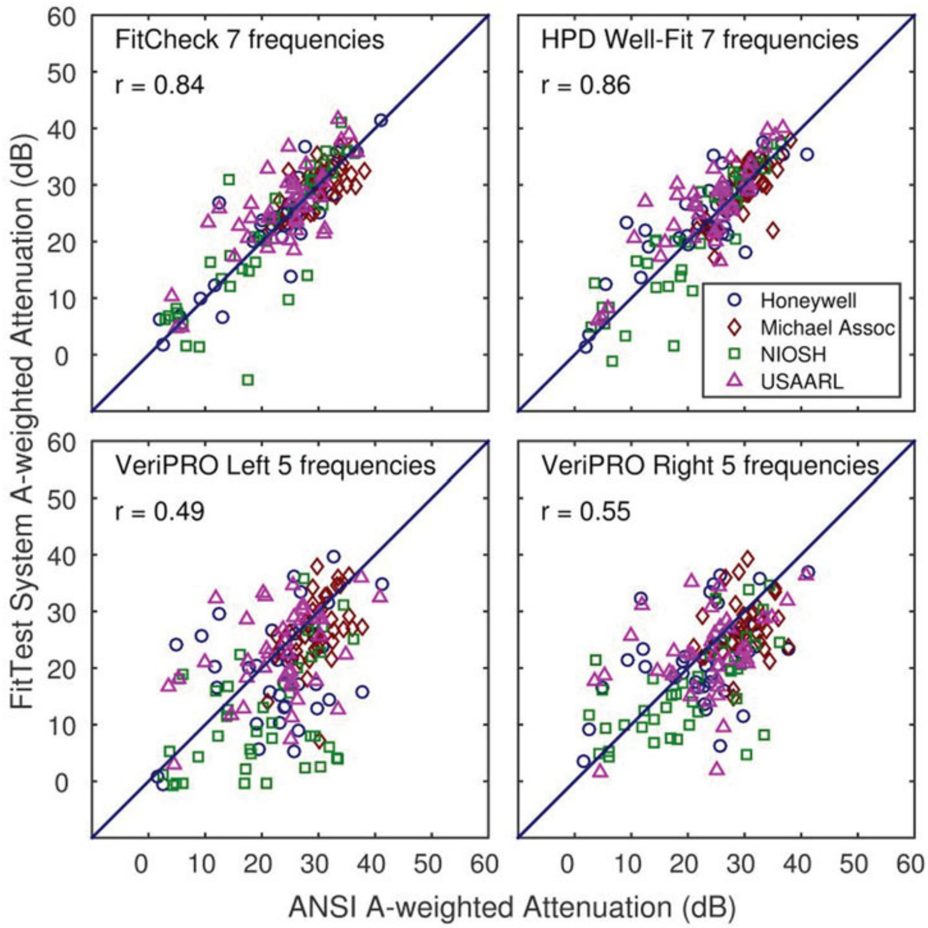

Figure 3. Comparison of fit-test system results to the corresponding seven-frequency ANSI laboratory results and the Pearson product-moment correlation coefficient (r). Different symbol shapes represent the four participating test laboratories: Honeywell (○), Michael & Associates (◊), NIOSH (□), and USAARL (△). The diagonal lines indicate the perfect correlation of 1.0.

more tightly grouped around the one-to-one correlation line compared to the subjects from the NIOSH lab.

Figure 4 depicts the average A-weighted differences among each fit-test system and the corresponding ANSI test results at each laboratory. The A-weighted attenuations were calculated using Equation (1) and the difference between the ANSI and fit-test results were determined for both trials of each subject's tests. Because we were interested in comparing the results against the A-weighted ANSI attenuation, the data from all seven test frequencies (125–8000 Hz) were used to compute the A-weighted attenuation for the ANSI, FitCheck, and HPD Well-Fit systems. A-weighted attenuations were estimated using five frequencies (250–4000 Hz) for the VeriPRO system. The effect of making the comparison using A-weighted attenuations derived from five instead of seven frequency bands did not materially change any of the significance tests for the labs and fit-test systems except for the FitCheck system measured in the USAARL laboratory. Ideally, all results would fall horizontally on the 0-dB difference line, indicating that there was no difference

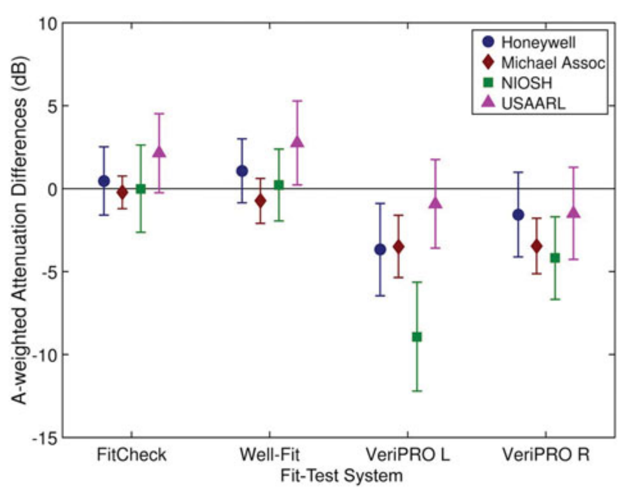

Figure 4. Mean differences between seven-frequency A-weighted ANSI attenuations and the fit-test system at each laboratory. Error bars represent 95% confidence intervals. For FitCheck and HPD Well-Fit, the differences are among the seven-frequency A-weighted attenuations. For the VeriPRO left and right, the differences are among the five-frequency A-weighted attenuations.

between the fit-test system results and the laboratory standard threshold-based method of attenuation testing.

Results for the FitCheck system were closest to the ANSI values at three of the four test locations. The USAARL lab's FitCheck results were not significantly different from zero ($p < 0.05$) with the seven-frequency A-weighted attenuation. For the HPD Well-Fit system, results from the Honeywell, Michael & Associates, and NIOSH labs were not significantly different from zero ($p < 0.05$) while the USAARL results were higher by 2–3 dB. For the VeriPRO system, the A-weighted attenuations were less than the ANSI A-weighted attenuation for both the right and left ears. The labs had a mixed level of statistically significant differences. For NIOSH and Michael & Associates, both the left and right ear differences were significant. At Honeywell, the left ear difference was significant but the right ear difference was not. For the USAARL lab, neither the right nor left ear differences were statistically significant. In general, it should be noted that the greater variability evident in Figure 1 within the Honeywell, NIOSH, and USAARL labs would have less likelihood of observing a statistically significant difference in the group data.

The statistical analysis yielded the overall decibel differences and 95% confidence intervals between the fit-test systems and the ANSI sound field method across the four testing laboratories: FitCheck = 0.6 (−0.5, 1.6); HPD Well-Fit = 0.8 (−0.2, 1.8); VeriPRO Left = −4.2 (−5.6, −2.9); VeriPRO Right = −2.7 (−3.9, −1.5) dB. The differences for FitCheck and HPD Well-Fit were not statistically significant at the ($p < 0.05$) level. VeriPRO exhibited statistically significant bias to underestimate the seven-frequency A-weighted attenuation by about 3–4 dB.

Discussion

The purpose of this study was to understand the accuracy and variability of fit-test systems relative to the ANSI method. From these analyses, the two REAT-based systems were accurate to within about 1 dB and exhibited confidence intervals of about 2 dB compared to the ANSI sound-field tests. The Awt_5 from the VeriPRO system underestimated Awt_7 of the ANSI method by 3–4 dB. Depending upon the laboratory subjects and tester, the error varied between −8 (NIOSH) and −1 dB (USAARL). The width of the confidence interval increased as the average attenuation in each laboratory decreased. In other words, lower attenuation results had higher variability.

The attenuations were determined using all of the data provided by the systems. The VeriPRO did not test 125 or 8000 Hz. In Equation (1), A-weighting reduces the influence of 125 Hz and 8000 Hz by 16.1 and 1.1 dB, respectively. These frequencies can be dropped because attenuation 125 Hz is highly correlated with 250 and 500 Hz and attenuation at 8000 Hz is correlated strongly with 4000 Hz. The distribution of the differences between the Awt_5 – Awt_7 was examined. A maximum deviation of −0.7 dB was observed. The distribution was skewed to the left and more than 90% of the differences were between −0.7 and 0.0 dB. In investigations of A-weighted attenuations computed with fewer than seven frequencies, the choice of frequencies can dramatically affect the accuracy of the PAR calculation.

The FitCheck and HPD Well-Fit systems exhibited good agreement with the ANSI results across subjects. This is not surprising, considering the similarity between these REAT-based methods. VeriPRO attenuations were consistently lower than those measured with the other fit-test systems, particularly at the higher frequencies. The correlations for the A-weighted attenuation data in Figure 3 are perhaps indicative of the increased difficulty of performing the loudness balance task. Detection of a stimulus in the sound field and headphone implementations of REAT has been demonstrated to have standard deviations on the order of a few decibels. However, Franks et al.[13] reported larger standard deviations of the order of 9–11 dB for a bone conduction loudness balance task. In this study the subjects were not necessarily selected for the ability to insert the earplugs and attain a high attenuation. They were required to demonstrate proficiency with the REAT procedure, which likely reduced the differences between the ANSI and the REAT-under-headphone fit-tests. Given the poorer correlations between VeriPRO and the ANSI results, several changes may be warranted to the test procedure. If the subjects are unfamiliar with the test paradigm, then training should be provided before conducting an attenuation test. This recommendation would hold for both FitCheck and HPD Well-Fit. The psychophysical paradigm to determine loudness balance could be modified to develop a psychometric function for loudness balance rather than relying upon a single indication of matched right-ear and left-ear levels. If the tonal stimulus exhibits harmonics at higher levels, then subjects may perceive the loudness differently for signals with more harmonic content than less harmonic content. Using noise bands for the loudness balance may change the outcome.

Trompette et al.[14] reported that VeriPRO attenuation values underestimated REAT and varied with the earplug tested. Brueck[15] reported lower mean values for the PAR attenuations by frequency compared to the manufacturer's reported REAT values. Kotarbińska and Rogowski[16] also reported that PAR values measured with VeriPRO were lower than the NRR ratings for a wide range of earplugs. However, a primary difference between the current study and these other studies was that the current study maintained the same earplug fit across both the laboratory REAT and fit-test system tests. Comparing the PAR for a mixture of subjects with varying levels of experience to the subject panels used to establish the NRR, SNR, or HML rating could be expected to yield lower attenuation values. Because subjects were tested with the same fit of the earplugs across all systems, the comparisons have greater significance upon the ability of the fit-test systems to estimate the attenuation of an earplug's fit.

Inexperience with the loudness balance paradigm could be an explanation for the variability seen in the VeriPRO system results. With the exception of some of the experienced subjects, the likelihood is high that none had ever performed a loudness balance task. Threshold detection tasks are the foundation of audiometric testing and are different from the cognitive judgment tasks required for loudness comparisons. Workers in a hearing conservation program gain training in the detection task whereas the loudness balance task likely is new to them. Anecdotally, the VeriPRO system is frequently used in its Quick Check mode, which tests only 500 Hz. If a company uses the Quick Check mode to identify a protector that works well for an employee, then switches to the Complete Check of five frequencies the consistency of the results might improve. Also, the individual administering the test must be vigilant not to accept bad data during the testing. When the loudness balance test is conducted in the Complete Check mode, the 500 Hz test is flagged if the two matched attenuations do not agree within 6 dB. Thus, the operator can be faulted for results that have poor repeatability or reliability.

Two prior comparisons of loudness balance and laboratory REAT attenuation values found lower attenuation values would be obtained with the loudness balance method, particularly at the low and high frequencies.[17,18] The procedural method in Hershkowitz and Levine involved remembering the apparent loudness between having the earmuffs on and the presentation of a stimulus without earmuffs. The ABLB method overcomes the "memory" factor because the comparisons are made between ears in various states of occlusion. Hershkowitz and Levine suggested that a 6-dB correction be applied to the data "until such time as the attenuation of ear protectors measured by the threshold method can be compared with a direct measure of effectiveness." As reported by Rudmose,[19] subsequent studies of loudness balance to compare minimum audible pressure (MAP) and minimum audible field (MAF) thresholds have identified potential artifacts that could have affected Hershkowitz and Levine's findings. Source location (near or far) and isolation of the subject's chair was from the floor are two such artifacts that could influence the differences observed in thresholds. The current study maintained the fit of the earplugs to make a direct comparison of a loudness balance with REAT and eliminated the problems encountered by Hershkowitz and Levine.[17] The Hershkowitz and Levine study used a free field stimulus (MAF) while the present study used headphones to present the signal (MAP). Presumably the headphones sealed around the pinna and would not change any physiological noise effect between the open-ear and occluded ear canal conditions. The subjects' ears were in an occluded condition by wearing the headphones, so significant changes in the physiological noise should not have played a significant role as it might have for a free field stimulus loudness balance test.

Rimmer and Ellenbrecker[20] proposed using a bone-conduction loudness balance (BCLB) and Franks et al.[14] tested the method with a sound field paradigm. The variance for the BCLB attenuations was greater than those observed for open-ear REAT in sound field. Variance due to the fit of the hearing protector is inevitably confounded with any test of occluded conditions. Schulz[7] reported that loudness balance estimates of attenuation were less variable than the ANSI laboratory method in an internal study, and more variable (standard deviations were 2 dB higher) in another contracted evaluation of the VeriPRO system. The latter findings are more in concert with those of the current study.

HPD Well-Fit tended to produce a lower average attenuation level than the other methods at 125 Hz. This difference might be attributable to the novelty of the Method of Adjustment paradigm used by HPD Well-Fit. Subjects did not receive an opportunity to practice using HPD Well-Fit prior to obtaining the first open-ear measurement at 125 Hz. Therefore, training with using the HPD Well-Fit system is necessary before performing any tests that may be relied upon. Alternatively, when testing with HPD Well-Fit, the testing paradigm could be adjusted to use a starting frequency that will be repeated later during the testing. Otherwise, the first trial on a frequency could be discarded before estimating the personal attenuation rating (PAR).

In the Michael & Associates laboratory, both the FitCheck and HPD Well-Fit systems underestimated the attenuation measured by the ANSI system at 8000 Hz, which exceeded 40 dB. The maximum presentation levels of the headphones were limited to 80 dB SPL by the HPD Well-Fit system and none of the subjects exceeded 70 dB SPL. At 70 dB SPL and 8000 Hz, the FitCheck headphones used with the HPD Well-Fit system have total harmonic distortion less than 2%.[21] Further research is needed to identify the source of this difference and whether it was unique to the Michael & Associates laboratory.

Reliability testing

The HPD Well-Fit system tests repeatability in the threshold measurement at each frequency. The subjects were required to identify threshold three times at each frequency for the occluded and open-ear conditions. The tolerance was set to 6 dB for the range of three consecutive threshold identifications. If the subject failed to meet that tolerance, they were tested until they met the criterion. All of the subjects completed the HPD Well-Fit tests and satisfied the threshold repeatability tests.

The FitCheck system employed the same criteria as the ANSI REAT systems. The Békésy method required that no ascending reversal be below a descending reversal (peaks above valleys) and that the range of reversals be less than 20 dB and greater than 3 dB. If the subject failed the first Békésy threshold sequence, the frequency was immediately retested. A failure on the second trial resulted in a non-response in the database. All of the subjects have responses in the FitCheck datasets.

The VeriPRO system performs the reliability check for only the first open-ear frequency (500 Hz) and does not test reliability at other open-ear frequencies. For the occluded tests, VeriPRO compares the attenuation against what laboratory attenuation levels for the product being tested. If attenuation is too low at the first occluded frequency (500 Hz), then a warning is displayed so that the occluded loudness match can be repeated for that frequency. The VeriPRO system relies upon the subject's matches at each frequency to determine the loudness matching level. If the test paradigm were to derive a loudness matching psychometric function from multiple presentations above and below the reference tone in both the occluded and open-ear conditions, then the accuracy of the test might be improved and reliability could be tested as the test proceeded.

Personalizing the approach toward earplug attenuation measurement effectively eliminates the need for the NRR and avoids the pitfalls associated with it. Because quantitative fit-test results are typically reported as a personal attenuation rating, comparison of the resultant PAR values for the various systems becomes important. This aspect of fit-testing is still under development. A document outlining the performance criteria for individual hearing protector fit-testing systems is expected to be released by Working Group 11 of the ANSI Accredited Standards Committee on Noise (ANSI S12/WG 11).

Although the EPA-labeled 27-dB NRR provides information for selecting a hearing protector, this study—and many others preceding it—demonstrates that laboratory performance ratings do not predict the attenuation achieved in field situations (see last column of Table 2).[3,5,13,22–26] Revising the NRR to use the ANSI/ASA S12.68 estimation methods would yield a more relevant rating number that reflects both the performance of untrained users and highly experienced and motivated users.[5,25,27] The proposed revisions to NRR labeling regulation address known problems with the current labeling system.[28] Some of the comments received by the EPA urged adoption of the ANSI/ASA S12.6 Method B inexperienced-subject fit protocol. Others proposed that the EPA should provide guidance on how to incorporate fit-testing into the rating process. In fact, the EPA's mandate is to provide the consumer of a hearing protection device information about the potential performance of the product for reducing noise.[2,28,29] Thus, the appropriate choice for EPA was to select the Method A protocol for a trained-subject fit for the proposed rule. Method A allows the manufacturer to understand the performance capability of a product, and the labeled rating would become the target for attenuation as measured by a fit-test system. The ANSI/ASA S12.68 rating methods produce a range of attenuation values that will facilitate hearing protector selection.[27] If and when the EPA promulgates an updated NRR labeling rule, a paradigm shift in the administration of hearing loss prevention programs may ensue.

Future research

Although the accepted ANSI laboratory method for REAT testing (ANSI/ASA S12.6-2008) prescribes a modified Békésy tracking paradigm, some of the newer commercially-available field methods for earplug fit-testing do not employ this method.[6] Therefore, research should include additional investigations into the different methods, with particular attention to any minimum practice requirements necessary to yield accurate results among the different systems.

In addition to overall accuracy and validity, issues such as timing of the different test methods also need to be researched. Minimizing employees' time away from work will help provide an incentive towards universal acceptance of fit testing in industrial workplaces.

Conclusions

As discussed elsewhere in the bourgeoning literature involving earplug fit-test systems, there are many benefits to the hearing conservation program administrator. A few of these include training of wearers in correct fitting procedures; random as well as routine field sampling of protector effectiveness; documentation that training was provided and that proper protection was provided to the employee; and identification of failing or deteriorating protectors and changes in ear physiology. Efforts are underway to fully understand the capabilities of these field measurement systems and promote their use in industrial hearing loss prevention programs.

Four independent laboratories each tested three hearing protector fit-test systems on groups of subjects wearing the same brand/model of earplug. Testing also was conducted according to the standard laboratory REAT procedure, which was used as the reference value for each individual's fit-test. The laboratories exhibited significantly different results; however, the focus of this study was to compare each of the fit-testing systems to the ANSI laboratory test. A-weighted attenuations measured with FitCheck and HPD Well-Fit systems demonstrated approximately ±2 dB agreement with the ANSI method, and were not statistically different from the ANSI laboratory method in any of the laboratories. A-weighted attenuations measured with VeriPRO tended to yield attenuations lower than the ANSI attenuation, and were significantly different ($p < 0.05$) from the ANSI laboratory results. None of the fit-test systems produced attenuation values that exceeded the laboratory results. Experience with earplug fitting and related attenuation measurements may be indicative of the variability of the fit-testing results. Training in the particular paradigm also is important before conducting individual fit-testing.

Acknowledgments

Claire Collard-Johnson collected data at the NIOSH laboratory as a part of her Capstone project while enrolled at the University of Cincinnati Department of Communication Sciences Disorders. Brad Witt and Theresa Schulz from Honeywell Safety Products are acknowledged for their thoughtful discussions of the VeriPRO system. Elliott Berger and Cameron Fackler from 3M Personal Safety Division contributed to the discussion of the findings of this study.

Disclaimer

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention, the National Institute for Occupational Safety and Health (NIOSH), or the U.S. Army Aeromedical Research Laboratories. Mention of any company or product does not constitute endorsement by NIOSH or the U.S. Army Aeromedical Research Laboratories.

References

[1] **Berger, E.H.:** Methods of measuring the attenuation of hearing protection devices. *J. Acoust. Soc. Am. 79*(6):1655–1687 (1986).

[2] "Noise Labeling Requirements for Hearing Protectors," *Code of Federal Regulations Title 40*, Part 211. 1979. pp. 56130–56147.

[3] **Berger, E.H., J.R. Franks, and F. Lindgren:** International review of field studies of hearing protector attenuation. In *Scientific Basis of Noise-Induced Hearing Loss*, A. Axlesson, H. Borchgrevink, R. P.Hamernik, P. Hellstrom, D. Henderson, and R. J. Salvi, (eds.). New York: Thieme Medical Publishers, Inc., 1996. pp. 361–377.

[4] **American National Standards Institute (ANSI):** *American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors* [ANSI/ASA S12.6–1997(R2002)]. [Standard] New York: ANSI, 2002.

[5] **Murphy, W.J., D.C. Byrne, D. Gauger, et al.:** "Results of the National Institute for Occupational Safety and Health - U.S. Environmental Protection Agency Interlaboratory Comparison of American National Standards Institute S12.6-1997 Methods A and B." *J. Acoust. Soc. Am. 125*(5):3262–3277 (2009).

[6] **Hager, L.D.:** Fit-testing hearing protectors: An idea whose time has come. *Noise Health 13*(51):147–151 (2011).

[7] **Schulz, T.Y.:** Individual fit-testing of hearing protectors: A review of uses. *Noise Health 13*(51):152–162 (2011).

[8] **NHCA - NIOSH - OSHA Alliance:** Best Practice Bulletin: Hearing Protection - Emerging Trends: Individual Fit Testing, 1–3 (2008).

[9] **National Institute for Occupational Safety and Health:** *A Real-Ear Field Method for the Measurement of the Noise Attenuation of Insert-type Hearing Protectors*, by **P.L. Michael, R.L. Kerlin, G.R. Bienvenue, J.H. Prout, and J.I. Shampan** (Publication No. 76-181). DHEW NIOSH, June 1976.

[10] **American National Standards Institute (ANSI):** *American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors* [ANSI/ASA S12.6-2008]. [Standard] New York: ANSI, 2008.

[11] **Soli, S., A. Vermiglio, and V. Larson:** "A system for assessing the fit of hearing protectors in the field." Presentation at the National Hearing Conservation Association Conference, Tucson, AZ, February 24–26, 2005.

[12] **Snedecor, G.W., and W. Cochran:** *Statistical Methods*. 6th ed. Ames, IA: University Press, 1967.

[13] **Franks, J.R., W.J. Murphy, D.A. Harris, J.L. Johnson, and P.B. Shaw:** Alternative field methods for measuring hearing protector performance. *AIHA J. 64*:501–509 (2003).

[14] **Trompette, N., A. Kusy, and J. Ducourneau:** Suitability of commercial systems for earplug individual fit testing. *Appl. Acoust. 90*:88–94 (2015).

[15] **Health and Safety Executive:** *Assessment of subjective and objective measurement systems of earplug attenuation on an individual*, by **L. Brueck**, HSE RR975, Derbyshire, UK, 2013.

[16] **Kotarbińska, E., and K. Rogowski:** Real-world performance of earplugs measured by Personal Attenuation Rating. In *Forum Acusticum*. pp. 1–8 (2014).

[17] **Hershkowitz, J., and L.M. Levine:** Attenuation of ear protectors by loudness balance and threshold methods. *J. Acoust. Soc. Am. 29*(8):889–894 (1957).

[18] **Weinreb, L., and M.L. Touger:** Variation in ear protector attenuation as measured by different methods. *J. Acoust. Soc. Am. 32*(2):245–224 (1960).

[19] **Rudmose, W.:** The case of the missing 6 dB. *J. Acoust. Soc. Am. 71*(3):650–659 (1982).

[20] **Rimmer, T.W., and M.J. Ellenbecker:** Hearing protector attenuation measurement by bone conduction loudness

balance compared with real-ear attenuation at threshold in a sound field. *Appl. Occup. Environ. Hyg. 12*:69–75 (1997).

[21] **Murphy, W.J., and J.R. Franks:** Evaluation of a FitCheck hearing protector test system. *J. Acoust. Soc. Am. 106*(4–2):2263 (1999).

[22] **Royster, J. D., E.H. Berger, C.J. Merry, et al.:** Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part I. Research of Working Group 11, Accredited Standards Committee S12, Noise, *J. Acoust. Soc. Am. 99*:1506–1526 (1996).

[23] **Berger, E.H., J.R. Franks, A. Behar, et al.:** Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part III. The validity of using subject-fit data. *J. Acoust. Soc. Am. 103*:665–72 (1998).

[24] **Murphy, W.J., J.R. Franks, E.H. Berger, et al.:** Development of a new standard laboratory protocol for estimation of the field attenuation of hearing protection devices: Sample size necessary to provide acceptable reproducibility. *J. Acoust. Soc. Am. 115*(1):331–323 (2004).

[25] **Murphy, W.J., M.R. Stephenson, D.C. Byrne, B. Witt, and J. Duran:** Effects of training on hearing protector attenuation. *Noise Health 13*:132–141 (2011).

[26] **Joseph, A., J. Punch, M.R. Stephenson, E. Wolfe, N. Paneth, and W.J. Murphy:** The Effects of training format on earplug performance. *Int. J Aud. 46*:609–618 (2007).

[27] **American National Standards Institute (ANSI):** *Methods of Estimating Effective A-weighted Sound Pressure Levels When Hearing Protectors are Worn*. [ANSI/ASA S12.68-2007 R2012]. [Standard] New York: ANSI, 2012.

[28] **EPA:** "40 CFR 211 Subpart B Proposed Revised Rule," *Federal Register 74*:149 (August 5, 2009). pp. 39150–39196.

[29] "Noise Control Act of 1972", Public Law 92-574, Section 8, *U.S. Code*, Title 42, Pt. 4907, Labeling. Oct. 27, 1972.