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

William J. Murphy, David C. Byrne, Dan Gauger, William A. Ahroon, Elliott Berger, Samir N. Y. Gerges, Richard McKinley, Brad Witt, and Edward F. Krieg

Citation: The Journal of the Acoustical Society of America **125**, 3262 (2009); doi: 10.1121/1.3095803

View online: https://doi.org/10.1121/1.3095803

View Table of Contents: http://asa.scitation.org/toc/jas/125/5

Published by the Acoustical Society of America

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

William J. Murphy[a] and David C. Byrne
Hearing Loss Prevention Team, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, MS C-27, Cincinnati, Ohio 45226-1998

Dan Gauger
Bose Corporation, MS271E, 145 Pennsylvania Avenue, Framingham, Massachusetts 01701-9168

William A. Ahroon
U.S. Army Aeromedical Research Laboratory, 6901 Andrews Avenue, P.O. Box 620577, Fort Rucker, Alabama 36362-0577

Elliott Berger
Aearo Technologies, 7911 Zionsville Road, Indianapolis, Indiana 46268-1657

Samir N. Y. Gerges
Departamento de Engenharia Mecânica, Laboratório de Vibrações e Acústica, (LARI and LAEPI), Universidade Federal de Santa Catarina, Cx. P. 476, CEP 88040-900, Florianópolis, SC, Brazil

Richard McKinley
Human Effectiveness Directorate, AFRL/HECB, 2610 7th Street, Building 441, Dayton, Ohio 45433

Brad Witt
Howard Leight Industries, 7828 Waterville Road, San Diego, California 92154

Edward F. Krieg
Division of Applied Research and Technology, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, MS C-22, Cincinnati, Ohio 45226

(Received 22 September 2008; revised 8 January 2009; accepted 9 February 2009)

The National Institute for Occupational Safety and Health and the Environmental Protection Agency sponsored the completion of an interlaboratory study to compare two fitting protocols specified by ANSI S12.6-1997 (R2002) [(2002). American National Standard Methods for the Measuring Real-Ear Attenuation of Hearing Protectors, American National Standards Institute, New York]. Six hearing protection devices (two earmuffs, foam, premolded, custom-molded earplugs, and canal-caps) were tested in six laboratories using the experimenter-supervised, Method A, and (naïve) subject-fit, Method B, protocols with 24 subjects per laboratory. Within-subject, between-subject, and between-laboratory standard deviations were determined for individual frequencies and *A*-weighted attenuations. The differences for the within-subject standard deviations were not statistically significant between Methods A and B. Using between-subject standard deviations from Method A, 3–12 subjects would be required to identify 6-dB differences between attenuation distributions. Whereas using between-subject standard deviations from Method B, 5–19 subjects would be required to identify 6-dB differences in attenuation distributions of a product tested within the same laboratory. However, the between-laboratory standard deviations for Method B were −0.1 to 3.0 dB less than the Method A results. These differences resulted in considerably more subjects being required to identify statistically significant differences between laboratories for Method A (12–132 subjects) than for Method B (9–28 subjects).
 [DOI: 10.1121/1.3095803]

PACS number(s): 43.66.Vt, 43.50.Hg, 43.66.Yw [BSF] Pages: 3262–3277

I. INTRODUCTION

In March 2003, the United States Environmental Protection Agency (EPA) conducted a workshop to bring together parties involved in the manufacture, sale, testing, and use of hearing protection devices (HPDs). These parties included government, academia, manufacturers, and testing laboratories. During this two-day workshop, possible changes in the EPA's regulation for hearing protector labeling, 40 CFR 211 subpart B (EPA, 1978), were discussed during lecture and facilitated discussions. One of the more controversial discussion topics was the choice of testing methods for measuring

[a]Electronic mail: wmurphy@cdc.gov

the attenuation of HPDs among those specified in the American National Standard Method for Measuring the Real-Ear Attenuation of Hearing Protectors (ANSI S12.6-1997 (R2002), 2002).

The two methods under consideration were an experimenter-supervised fit (Method A) and a naïve subject-fit (Method B). Method A allows the use of subjects who have received training or have experience fitting HPDs. However, Method B requires subjects who have not received previous one-on-one training and have limited experience with protector testing, with wearing HPDs, and with computer-based or video training in the correct fitting of HPDs. The workshop participants identified the need for a direct comparison of the two methods as a priority before revising the regulation. For insert earplugs, the role of the experimenter has proven to be a significant factor in the amount of attenuation achieved by a particular laboratory. According to Berger *et al.* (1998), attenuations from naïve subjects were representative of the upper quartile of the real-world attenuation measurements. Under Method B, naïve subjects are recruited to assess the performance expected from real-world users.

Interlaboratory studies of this scope are few. In the early 1990s, two studies of informed-user fit and subject-fit protocols were completed. The results of the first study were unpublished and motivated the design of the second study in 1992. Four papers resulted from the second study: Royster *et al.* (1996), Berger *et al.* (1998), Murphy *et al.* (2002), and Murphy *et al.* (2004). As a consequence of the interlaboratory study, ANSI S12.6-1984 was revised in 1997 and included the Method B subject-fit test real-ear attenuation at threshold (REAT). In the second study, as documented in Royster *et al.* (1996), four hearing protectors were evaluated: the Bilsom Blue earmuff, the E·A·R® Classic® foam earplug, the Allsafe V-51R single-flanged premolded earplug, and the Willson EP100 triple-flanged earplug. The devices were selected for laboratory attenuation testing based on the availability of real-world studies for comparison. Berger *et al.* (1998) examined the ability to predict real-world protector performance of the naïve subject-fit protocol as compared to the experimenter-fit protocol on which the current US Noise Reduction Rating (NRR) is based. They concluded that the naïve subject-fit test corresponded more closely, approximating the upper quartile of the real-world studies. Similarly, Berger *et al.* (1998) showed that the naïve subject-fit test rank-ordered relative protector performance almost identically to the real-world studies. Murphy *et al.* (2002) developed mathematical models to describe the distribution of attenuations as a function of frequency and protector. The attenuation distributions from premolded earplugs tended to be bimodal at frequencies below 1000 Hz due to poorly-fit earplugs for some subjects. The distributions were accurately modeled for all frequencies and protectors by a mixed Gaussian distribution. For other protectors and frequencies, the data were normally distributed and did not require a mixed distribution. Murphy *et al.* (2004) reported the statistical analysis of the within-subject, between-subject, and between-laboratory variabilities necessary to estimate sample sizes to determine repeatability and reproducibility. Repeatability characterizes the expected variability if the protectors were to be tested in the same laboratory with the same subject panel. Reproducibility characterizes the expected variability if the protectors were to be tested with a different panel of subjects in the same laboratory or in a different laboratory. This analysis formalized the statistical justification for the standard's use of 10 subjects for earmuffs and 20 subjects for earplugs and semi-aural insert HPDs.

In 2004, European acoustic research and HPD testing laboratories reported a round-robin study where subjects were tested with the same products in several laboratories (Poulsen and Hagerman, 2004). The Nordic round-robin found no significant differences among the laboratories for the several devices tested. In Royster *et al.* (1996), the results from four studies reported between 1976 and 1986 were summarized, finding low variability for repeated measures of subjects' attenuations, large intersubject variability, and larger interlaboratory variability. Beyond this limited set of studies, no additional studies have significant bearing on the results to be presented here.

In November 2004, a test protocol was prepared for the current study. Six testing laboratories agreed to participate: Aearo/E·A·R® E·A·RCAL^{SM} test laboratory (E·A·RCAL), Howard Leight Industries (HLI) test laboratory, Brazil's Laboratory for Acoustic Research Institute (LARI), the National Institute for Occupational Safety and Health Robert A. Taft Laboratory (NIOSH), the U.S. Army Aeromedical Research Laboratory (USAARL), and the U.S. Air Force Research Laboratory (AFRL). The test protocol specified six products to be tested by each laboratory with 24 naïve subjects recruited from the respective local communities, for a total of 144 subjects. Subjects were tested first according to ANSI S12.6-1997 (R2002) (2002) Method B then, after individual instruction, according to Method A. The laboratories began testing in January 2005 and all tests were completed by August 2006.

The subjects' mean hearing thresholds, anthropometric information, mean attenuations, and statistical analyses of the results for repeatability and reproducibility are given in this manuscript.

II. METHODS

A. ANSI S12.6-1997 (R2002)

ANSI S12.6 specifies two protocols to assess the REAT for a hearing protection device: Method A experimenter-supervised fit and Method B subject fit. The experimental protocol required each laboratory to recruit naïve subjects (i.e., no prior experience with testing and limited experience in using hearing protection devices) for the Method B testing first. Each subject was then trained in the fitting of the different protectors and tested according to the Method-A protocol. The protocol deviated from S12.6 in that the standard allows continued use of experienced subjects for Method A whereas only newly-trained naïve subjects were used in this experiment. The participating laboratories decided *post hoc* that this difference had small effect on the reproducibility conclusions reached. The Method-A protocol allows the experimenter to instruct the subject with any variety of training

TABLE I. Subject recruitment and retention by laboratory. Participating laboratories were AFRL, E·A·RCAL, HLI, LARI, NIOSH, and USAARL

Laboratory	Subjects recruited	Subjects tested	Subjects rejected	Subjects dropped
AFRL	36	24	5	7
E·A·RCAL	47	24	17	6
HLI	30	24	6	0
LARI	37	24	13	0
NIOSH	27	24	3	0
USAARL	30	24	0	6

TABLE III. Means and standard deviations of ear canal size (diameters), bitragion width, and head height.

Lab	Right canal size (cm)	Left canal size (cm)	Bitragion width (cm)	Head height (cm)
AFRL	0.88 ± 0.10	0.87 ± 0.09	13.48 ± 0.64	12.75 ± 0.74
EARCal	0.89 ± 0.10	0.89 ± 0.10	14.17 ± 0.73	12.34 ± 1.04
HLI	1.01 ± 0.09	1.01 ± 0.09	13.44 ± 1.62	13.79 ± 0.99
LARI	0.95 ± 0.08	0.95 ± 0.08	13.39 ± 0.85	12.38 ± 0.95
NIOSH	0.95 ± 0.10	0.93 ± 0.10	13.83 ± 0.81	13.97 ± 1.20
USAARL	0.94 ± 0.12	0.94 ± 0.12	13.83 ± 1.31	14.21 ± 1.35

materials and personal demonstration. However, as specified by Method A, the experimenter was not allowed to fit the protector on the subject, though if the fit were judged to be inadequate the experimenter could instruct the subject to refit the product. Laboratories followed their normal practice for instruction; no attempt was made *a priori* to standardize this aspect of the test.

Subjects were informed of any potential risk that they might face during the testing in the laboratory. The human subject use protocols were approved by the CDC-NIOSH human subjects review board, were reviewed by local review boards at USAARL and WPAFB, and complied with the ethical principles of the Acoustical Society of America.

B. Subject selection

Within each of the six laboratories, 24 adult subjects (12 females, 12 males) were recruited. In Table I, the statistics for the number of subjects who were recruited, tested, rejected, or dropped from the study are reported. As specified by ANSI S12.6, the subjects were prohibited from having received prior one-on-one training, were allowed only limited experience with hearing protector testing, were prohibited from having received computer-based or video training in the use of hearing protectors, and were allowed limited experience in wearing hearing protectors during the previous 2 year period. In addition, subjects were required to have normal anatomy of the external ear and ear canal (i.e., no obvious physical deformities), normal otoscopy, and hearing thresholds better than 25 dB HL (re ANSI S3.6-1996) at all test frequencies (125, 250, 500, 1000, 2000, 4000, 8000 Hz). The hearing thresholds were measured using standard audiometric procedures.

The ear canal size, bitragion width (width of the head at the tragus), and head height (tragus to crown) were measured. The ear canal size was measured according to ANSI S12.6 Annex D using the EARGAGE™ which has five different diameters: 7.62, 8.48, 9.27, 10.46, and 11.53 mm. The hearing thresholds and standard deviations are reported in Table II. The anthropometric results are reported in Table III.

C. Products under test

All product samples were purchased on the open market and were provided to the participating laboratories by the NIOSH organizers. Two earmuffs were selected: the Peltor Tactical-Pro and the 3M 1427. The Tactical-Pro earmuff is a sound-restoration electronic muff, which is intended to be worn with the headband over the head. Testing was conducted with batteries in the battery compartment; however, the electronics were turned off. The headband for the 3M 1427 passive earmuff can be worn in three different positions: over the head, under the chin, and behind the head. In this study, the headband was worn behind the head with the

TABLE II. Mean hearing threshold levels and standard deviations (dB HL) for right and left ears.

Laboratory	Ear	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
AFRL	Right	8.8 ± 4.9	3.3 ± 4.1	5.0 ± 3.9	4.2 ± 4.3	2.5 ± 3.6	4.0 ± 4.9	7.9 ± 8.2
	Left	8.5 ± 6.2	3.3 ± 3.5	4.2 ± 4.3	4.4 ± 4.0	2.9 ± 4.1	5.6 ± 6.0	4.8 ± 4.8
EARCal	Right	5.2 ± 6.8	3.9 ± 7.1	2.1 ± 6.1	5.0 ± 5.7	5.4 ± 6.1	10.2 ± 7.1	4.1 ± 7.0
	Left	6.6 ± 7.2	4.4 ± 7.5	2.0 ± 7.9	5.8 ± 6.3	6.2 ± 7.5	7.7 ± 6.9	4.7 ± 7.6
HLI	Right	7.1 ± 6.2	6.5 ± 5.2	3.5 ± 5.2	4.2 ± 5.2	2.9 ± 3.9	3.8 ± 4.9	6.7 ± 4.8
	Left	6.3 ± 6.8	5.0 ± 4.7	2.7 ± 3.9	3.1 ± 3.8	3.1 ± 4.4	3.5 ± 5.6	12.3 ± 8.5
LARI	Right	10.2 ± 4.5	8.1 ± 3.6	7.9 ± 4.4	7.5 ± 5.1	6.0 ± 4.7	8.1 ± 7.2	9.6 ± 5.7
	Left	11.7 ± 4.8	10.0 ± 4.2	9.4 ± 4.3	6.0 ± 4.4	5.4 ± 4.4	7.1 ± 5.3	7.7 ± 4.7
NIOSH	Right	6.7 ± 5.2	3.8 ± 4.2	5.2 ± 4.3	1.5 ± 3.5	1.0 ± 5.7	3.3 ± 7.9	6.5 ± 9.0
	Left	5.0 ± 5.3	3.1 ± 5.7	4.4 ± 5.4	2.7 ± 5.5	2.9 ± 7.9	4.2 ± 7.2	7.1 ± 9.1
USAARL	Right	9.8 ± 5.0	8.5 ± 4.8	6.9 ± 5.1	3.3 ± 4.3	0.4 ± 3.9	5.2 ± 7.6	7.3 ± 6.4
	Left	9.8 ± 4.8	9.0 ± 4.7	6.7 ± 5.5	4.6 ± 5.9	2.7 ± 4.2	4.0 ± 5.1	7.5 ± 4.4

TABLE IV. Means and standard deviations of headband clamping force measured in Newtons. (*N* is the number of hearing protectors.)

		Peltor TacticalPro		3M 1427		Moldex Jazzband	
Lab	Force system	Force	*N*	Force	*N*	Force	*N*
AFRL	INSPEC	11.9 ± 0.5	3	11.9 ± 0.5	3	2.5 ± 0.1	24
EARCal	INSPEC	10.2 ± 0.3	3	10.5 ± 0.3	3	2.5 ± 0.1	24
HLI	Load Cell	11.9 ± 0.5	3	10.9 ± 0.6	3	2.3 ± 0.1	24
LARI	Force Gauge	10.2 ± 0.2	3	11.1 ± 0.6	3	3.6 ± 0.2	24
NIOSH	Michael Assoc.	11.2 ± 0.1	3	10.5 ± 0.3	3	2.7 ± 0.1	24
USAARL	Force Gauge	10.5 ± 0.9	3	10.5 ± 0.8	3	2.5 ± 0.5	24

adjustable crown strap placed over the head to prevent the muff from sliding downward. The two muffs were chosen to represent an over-the-head and a behind-the-head earmuff. Twenty pairs of each earmuff were purchased with each laboratory receiving three pairs to be tested in a balanced manner across the 24 test subjects. The two remaining pairs were kept as spares in the event that parts required replacement. Only the vinyl crown strap for the 1427 needed replacement for three of the earmuffs during the course of the study.

Four earplugs were selected: E·A·R® Classic® foam plugs, Howard Leight AirSoft premolded plug, Custom Protect Ear dB Blocker custom silicon earplug, and the Moldex-Metric JazzBand banded hearing protector. The Classic was included to provide reference with the previous studies. The AirSoft was selected because it is a flanged premolded protector. One box of premolded earplugs was provided to each laboratory. The dB Blocker was selected to examine the variability of a custom-molded product across laboratories. Each subject's earmold impression was collected and one pair of custom-molded earplugs were manufactured for each test subject. The Moldex-Metric JazzBand, a banded hearing protector, achieves a seal at the entrance to the ear canal. Twenty-four JazzBands were provided to each laboratory with sufficient replacement tips.

The headband force of each earmuff and canal-cap protector was measured by the participating laboratories. AFRL and E·A·RCAL used the commercially available INSPEC system. NIOSH used the Michaels and Associates headband force system. The other laboratories, Howard Leight, LARI, and USAARL, utilized either a force gauge or a load cell as a part of a custom-built system. Means and standard deviations of headband force were reported in Table IV. The results were not statistically different across laboratories. Although AFRL measured the highest headband force for the earmuffs, it did not measure the highest force for the JazzBand products. Increased headband force can improve attenuation; however, the differences measured here were uncorrelated with the attenuation measurements to be discussed later.

D. Test procedure

Each subject was trained to perform the threshold test in the diffuse sound field at the respective laboratory. Subjects were required to produce three open-ear (unoccluded) thresholds that had a range of no more than 5 dB. Product testing was counter-balanced for product order and occluded/unoccluded order. A subject was required to complete the tests in occluded/unoccluded pairs. All of the Method-B tests for a subject were completed before the subject proceeded with any Method-A tests. Depending on the laboratory, each subject spent between five to seven visits to complete qualification and all product tests.

III. RESULTS

A. Attenuations by frequency

For each laboratory, device, and frequency, the mean Method-A attenuations are given in Table V and the mean Method-B attenuations are given in Table VI. The Method-B attenuations were generally less than the Method-A attenuations measured after the subjects had received instruction and training. For the earmuffs, the low-frequency attenuations ranged from about 10 to 20 dB at 125 and 250 Hz and from about 20 to 30 dB at 500 Hz. At high frequencies (1000–8000 Hz), the attenuations were about 30–40 dB. The Method-A earmuff attenuations were nearly the same or were a few decibels greater than the Method-B earmuff attenuations at all frequencies. For the earplugs tested under Method A, the protectors exhibited about 15–30 dB of attenuation in the low frequencies, 125–250 Hz and ranged from about 25–45 dB at the higher frequencies (500–8000 Hz).

B. *A*-weighted attenuation

Gauger and Berger (2004) conducted a comprehensive analysis of hearing protector rating methods and found that the use of an *A*-weighted statistic provided more accurate estimates of the effective exposure level when hearing protectors are worn. Current methods such as the NRR or the single number rating (SNR) are *C*-weighted statistics in that they are designed to be subtracted from a *C*-weighted exposure level. Following Gauger Berger's (2004) analysis, the Acoustical Society of America's accredited standards committee on noise, Working Group 11 developed ANSI S12.68-2007 that estimates the attenuation for each subject in a test panel across a population of noises (ANSI S12.68, 2007; Gauger and Berger, 2004; Johnson and Nixon, 1974; Kroes *et al.*, 1975). For each noise, the *A*-weighted attenuation is computed and the average and standard deviation of the at-

TABLE V. Method-A mean attenuations and standard deviations in dB at each test frequency, by laboratory and protector.

Lab	Protector	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
AFRL	TacPro	10.8 ± 2.5	18.2 ± 2.3	26.8 ± 2.7	35.1 ± 2.0	34.6 ± 2.5	35.4 ± 2.7	38.1 ± 2.6
EARCAL	TacPro	14.2 ± 5.2	19.2 ± 6.1	28.6 ± 6.9	32.9 ± 5.3	32.9 ± 4.7	35.0 ± 8.0	36.1 ± 6.4
HLI	TacPro	15.0 ± 2.7	21.9 ± 2.4	32.0 ± 2.6	34.2 ± 2.8	33.7 ± 3.1	37.9 ± 2.6	40.8 ± 2.1
LARI	TacPro	11.4 ± 3.9	20.3 ± 5.4	31.4 ± 2.2	32.6 ± 2.9	34.2 ± 1.9	35.3 ± 4.5	36.2 ± 2.9
NIOSH	TacPro	15.0 ± 3.8	21.4 ± 3.8	31.1 ± 3.4	32.3 ± 4.3	35.1 ± 3.2	38.1 ± 2.4	40.6 ± 3.8
USAARL	TacPro	12.6 ± 3.7	19.5 ± 3.7	24.9 ± 4.0	28.2 ± 3.8	28.9 ± 4.5	30.4 ± 4.8	35.0 ± 4.7
AFRL	1427	9.4 ± 6.6	13.4 ± 6.1	21.3 ± 5.8	36.5 ± 6.5	35.6 ± 2.7	34.5 ± 4.7	33.7 ± 4.6
EARCAL	1427	15.5 ± 7.3	16.3 ± 7.0	25.9 ± 6.6	36.0 ± 5.3	33.5 ± 3.3	36.2 ± 6.9	33.7 ± 6.8
HLI	1427	18.1 ± 4.0	20.3 ± 2.9	28.9 ± 4.3	36.0 ± 5.1	35.2 ± 2.7	38.6 ± 3.1	38.7 ± 2.4
LARI	1427	11.4 ± 5.8	16.3 ± 4.9	24.0 ± 6.1	33.1 ± 5.8	34.7 ± 3.0	32.3 ± 6.8	33.6 ± 4.3
NIOSH	1427	13.5 ± 8.2	16.8 ± 8.2	24.9 ± 6.8	32.1 ± 7.4	34.2 ± 4.3	36.6 ± 6.9	35.9 ± 5.1
USAARL	1427	13.6 ± 5.9	16.7 ± 6.4	22.5 ± 7.0	31.3 ± 6.7	29.8 ± 4.1	29.0 ± 6.8	30.9 ± 5.0
AFRL	dB Blocker	25.6 ± 6.2	25.7 ± 6.9	26.5 ± 6.0	30.7 ± 5.9	36.9 ± 4.5	41.0 ± 3.9	44.4 ± 4.2
EARCAL	dB Blocker	28.5 ± 9.0	27.4 ± 8.9	28.4 ± 9.7	27.8 ± 7.7	32.6 ± 5.4	39.9 ± 7.1	40.6 ± 6.6
HLI	dB Blocker	25.4 ± 6.5	24.9 ± 5.6	25.5 ± 5.2	25.5 ± 5.5	33.2 ± 4.5	40.8 ± 4.8	40.5 ± 7.1
LARI	dB Blocker	24.6 ± 9.1	27.1 ± 9.4	29.1 ± 8.6	28.4 ± 7.2	32.7 ± 6.8	39.7 ± 7.0	39.4 ± 6.3
NIOSH	dB Blocker	25.7 ± 7.8	27.2 ± 7.6	27.6 ± 7.1	25.7 ± 5.8	32.6 ± 4.2	40.0 ± 5.0	40.5 ± 7.0
USAARL	dB Blocker	24.9 ± 9.8	24.3 ± 8.4	24.4 ± 8.3	23.8 ± 5.9	27.7 ± 6.2	30.4 ± 6.7	34.2 ± 10.3
AFRL	JazzBand	19.0 ± 5.2	17.3 ± 5.0	16.6 ± 5.0	23.6 ± 3.9	32.6 ± 3.8	35.8 ± 3.3	38.3 ± 6.3
EARCAL	JazzBand	16.6 ± 9.1	15.5 ± 8.5	14.8 ± 7.6	18.0 ± 6.5	27.4 ± 5.7	34.9 ± 6.3	33.6 ± 7.6
HLI	JazzBand	24.7 ± 6.0	22.8 ± 5.5	21.4 ± 5.0	22.7 ± 4.4	32.3 ± 4.0	40.0 ± 4.3	42.9 ± 4.0
LARI	JazzBand	15.6 ± 8.7	15.0 ± 7.6	17.1 ± 6.2	18.4 ± 5.3	27.8 ± 5.9	34.8 ± 6.2	32.6 ± 6.8
NIOSH	JazzBand	20.9 ± 6.7	20.7 ± 6.7	18.4 ± 4.9	20.2 ± 4.9	30.3 ± 5.5	37.7 ± 4.4	37.0 ± 8.3
USAARL	JazzBand	19.1 ± 5.9	19.9 ± 5.6	17.4 ± 5.3	19.1 ± 5.7	26.8 ± 5.5	31.4 ± 6.1	33.5 ± 8.8
AFRL	Classic	30.9 ± 5.0	33.1 ± 4.7	35.4 ± 4.4	38.3 ± 3.6	36.9 ± 3.1	41.0 ± 2.2	47.3 ± 2.4
EARCAL	Classic	27.5 ± 8.9	28.5 ± 8.9	32.6 ± 9.7	31.0 ± 7.0	32.0 ± 4.1	41.2 ± 4.0	43.0 ± 5.0
HLI	Classic	32.1 ± 8.3	33.9 ± 7.6	36.9 ± 7.7	33.4 ± 6.0	34.2 ± 4.0	43.4 ± 3.3	46.3 ± 4.6
LARI	Classic	22.9 ± 4.9	26.6 ± 5.3	29.6 ± 5.9	28.7 ± 5.0	32.0 ± 3.3	40.7 ± 2.8	41.1 ± 4.8
NIOSH	Classic	24.4 ± 8.2	27.9 ± 8.7	31.0 ± 9.4	28.3 ± 7.0	33.0 ± 4.3	40.5 ± 3.2	43.5 ± 6.9
USAARL	Classic	19.7 ± 5.0	19.8 ± 4.4	19.6 ± 5.0	19.4 ± 4.9	26.9 ± 5.3	32.9 ± 5.3	35.5 ± 5.5
AFRL	AirSoft	18.8 ± 7.9	19.5 ± 7.6	20.5 ± 7.4	26.5 ± 6.5	32.0 ± 5.6	32.6 ± 7.5	40.0 ± 9.1
EARCAL	AirSoft	26.6 ± 7.5	25.8 ± 8.7	27.7 ± 9.4	27.1 ± 7.9	29.4 ± 4.7	34.9 ± 7.5	39.5 ± 9.3
HLI	AirSoft	29.4 ± 7.1	29.4 ± 6.6	31.3 ± 7.5	31.0 ± 5.6	32.0 ± 4.0	37.2 ± 6.9	44.7 ± 5.5
LARI	AirSoft	23.6 ± 8.8	25.2 ± 8.7	29.0 ± 9.7	27.8 ± 8.8	32.9 ± 5.0	37.7 ± 8.7	38.9 ± 7.8
NIOSH	AirSoft	24.9 ± 8.1	25.2 ± 8.1	27.3 ± 9.4	25.2 ± 8.1	30.3 ± 6.6	35.1 ± 8.9	40.6 ± 9.7
USAARL	AirSoft	20.3 ± 11.2	20.5 ± 10.4	20.9 ± 11.4	21.2 ± 10.7	26.1 ± 7.8	28.1 ± 8.8	33.1 ± 11.7

tenuations are used to estimate the SNR. To simplify the analysis, the *A*-weighted attenuation for pink noise is used to perform all of the subsequent analyses. The difference between the levels for *C*-weighted and *A*-weighted pink noise, $L_C - L_A$, is about 1.0 dB, which is close to the median difference, 1.8 dB, for the NIOSH 100 noises. Thus the results can be related to the expected results when the S12.68 standard is applied to determine a rating for the protector. The *A*-weighted attenuation using pink noise removes the additional element of variance across noises and allows more direct comparison of the attenuations due to the subject panels as well as retrospective comparison to the previous interlaboratory studies (Murphy *et al.*, 2004).

The overall *A*-weighted attenuation for each subject/protector combination was calculated using the following equation:

$$\text{Atten} = 10\log\left(\sum_{f=125}^{8000} 10^{L_f + A_f}\right) - 10\log\left(\sum_{f=125}^{8000} 10^{L_f + A_f - \text{Atten}_{f,\text{avg}}}\right), \quad (1)$$

where the test frequencies were $f = 125, 250, 500, 1000, 2000, 4000, 8000$ Hz. The *A*-weighting correction factors were $A_f = -16.1, -8.6, -3.2, 0.0, 1.2, 1.0, -1.1$ at the respective test frequencies. The noise spectrum levels were pink noise, $L_f = 100$ at all frequencies. The first summation yields 107.0 dB rounded to a tenth of a decibel. The attenuation, $\text{Atten}_{f,\text{avg}}$, measured from each subject's two paired occluded and unoccluded trials were averaged at each frequency. For example, rounded to a tenth of a decibel, one subject's Method-A attenuations were

TABLE VI. Method-B mean attenuations and standard deviations in dB at each test frequency, by laboratory and protector.

Lab	Protector	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
AFRL	TacPro	8.7 ± 3.2	16.5 ± 4.6	24.4 ± 4.1	34.4 ± 2.1	33.7 ± 2.6	34.5 ± 3.9	37.4 ± 4.3
EARCAL	TacPro	13.6 ± 5.9	19.2 ± 7.4	28.7 ± 9.1	32.9 ± 6.9	32.2 ± 5.7	35.1 ± 9.2	35.5 ± 7.8
HLI	TacPro	13.0 ± 5.1	19.2 ± 6.3	30.0 ± 5.6	33.1 ± 4.4	33.2 ± 2.9	36.0 ± 4.3	38.6 ± 3.5
LARI	TacPro	10.5 ± 4.2	20.3 ± 5.8	30.1 ± 2.7	31.9 ± 3.9	33.7 ± 3.2	34.2 ± 4.2	36.0 ± 5.1
NIOSH	TacPro	14.1 ± 4.8	21.1 ± 4.0	29.5 ± 5.0	31.3 ± 4.4	34.4 ± 3.7	36.5 ± 4.4	39.8 ± 3.8
USAARL	TacPro	12.7 ± 4.1	18.9 ± 4.0	25.0 ± 5.0	28.0 ± 4.4	29.3 ± 4.0	30.7 ± 5.5	34.0 ± 5.0
AFRL	1427	10.3 ± 4.0	14.9 ± 3.6	22.8 ± 4.2	36.9 ± 3.8	36.0 ± 3.1	34.7 ± 4.7	34.1 ± 5.0
EARCAL	1427	12.7 ± 6.8	16.1 ± 6.5	25.9 ± 6.2	34.9 ± 6.5	32.6 ± 4.1	33.5 ± 6.9	31.6 ± 7.0
HLI	1427	14.9 ± 7.4	18.0 ± 7.3	26.9 ± 6.7	34.6 ± 6.3	34.5 ± 3.0	35.4 ± 4.7	34.6 ± 5.2
LARI	1427	9.6 ± 5.6	14.5 ± 6.9	24.7 ± 5.5	33.4 ± 5.5	34.2 ± 3.7	31.1 ± 7.2	32.2 ± 5.9
NIOSH	1427	12.3 ± 7.6	15.9 ± 7.2	24.7 ± 7.3	33.1 ± 7.5	34.3 ± 4.9	34.6 ± 6.9	32.7 ± 6.7
USAARL	1427	13.6 ± 4.9	17.7 ± 5.6	23.1 ± 7.0	31.2 ± 5.6	28.9 ± 5.7	28.2 ± 7.4	27.8 ± 6.8
AFRL	dB Blocker	23.2 ± 9.0	24.5 ± 8.5	25.2 ± 9.2	29.2 ± 8.0	34.3 ± 8.1	38.7 ± 7.2	41.4 ± 8.8
EARCAL	dB Blocker	25.1 ± 13.0	25.9 ± 14.0	25.8 ± 14.1	24.8 ± 12.5	27.7 ± 11.7	34.9 ± 13.1	34.3 ± 15.1
HLI	dB Blocker	21.1 ± 7.8	21.7 ± 7.7	23.6 ± 7.8	23.3 ± 5.7	30.7 ± 4.9	39.4 ± 5.7	38.7 ± 6.1
LARI	dB Blocker	22.4 ± 10.4	24.1 ± 10.9	25.8 ± 10.4	25.5 ± 8.5	30.9 ± 8.5	35.5 ± 8.4	35.3 ± 9.9
NIOSH	dB Blocker	24.5 ± 9.8	25.4 ± 10.0	25.0 ± 10.4	23.1 ± 8.1	30.2 ± 7.2	37.2 ± 7.4	35.4 ± 9.9
USAARL	dB Blocker	21.0 ± 9.4	21.0 ± 10.3	21.7 ± 9.5	21.7 ± 9.2	27.0 ± 8.7	29.3 ± 9.0	31.9 ± 11.5
AFRL	JazzBand	16.4 ± 8.1	15.1 ± 7.9	14.6 ± 7.0	21.6 ± 5.6	29.7 ± 6.5	33.1 ± 6.5	33.4 ± 10.1
EARCAL	JazzBand	11.2 ± 9.3	11.3 ± 8.2	11.5 ± 8.0	14.8 ± 6.9	24.1 ± 7.4	31.2 ± 6.3	28.6 ± 8.6
HLI	JazzBand	17.6 ± 8.6	17.5 ± 8.1	17.3 ± 6.1	18.5 ± 6.0	27.1 ± 6.8	35.0 ± 7.2	33.0 ± 8.8
LARI	JazzBand	13.9 ± 9.0	13.7 ± 8.9	13.8 ± 9.7	14.6 ± 9.3	23.9 ± 10.0	30.4 ± 10.2	29.4 ± 11.7
NIOSH	JazzBand	18.3 ± 8.1	18.5 ± 7.4	17.3 ± 6.5	18.4 ± 5.2	27.8 ± 5.9	33.2 ± 6.2	33.4 ± 8.3
USAARL	JazzBand	17.0 ± 7.5	15.7 ± 7.3	15.2 ± 6.8	16.2 ± 7.2	23.6 ± 7.5	28.1 ± 9.3	30.0 ± 10.8
AFRL	Classic	18.5 ± 7.0	19.0 ± 6.8	21.2 ± 7.2	25.6 ± 5.7	31.7 ± 4.2	38.0 ± 4.1	40.9 ± 7.2
EARCAL	Classic	18.5 ± 9.8	19.0 ± 9.8	20.6 ± 11.8	20.8 ± 9.9	28.8 ± 6.0	37.1 ± 7.5	34.5 ± 10.7
HLI	Classic	19.9 ± 6.0	21.1 ± 6.3	23.1 ± 7.2	21.9 ± 6.3	29.3 ± 4.2	39.5 ± 5.1	40.2 ± 6.1
LARI	Classic	16.6 ± 6.5	19.0 ± 6.5	22.6 ± 7.3	21.5 ± 5.2	29.6 ± 3.8	37.8 ± 3.7	36.9 ± 5.2
NIOSH	Classic	17.3 ± 5.7	18.1 ± 5.2	19.5 ± 5.8	19.1 ± 4.5	27.9 ± 5.8	36.5 ± 4.8	34.6 ± 7.4
USAARL	Classic	16.6 ± 5.9	17.5 ± 5.5	17.7 ± 5.8	17.5 ± 6.0	25.4 ± 5.9	31.1 ± 7.6	31.5 ± 7.5
AFRL	AirSoft	16.2 ± 8.4	16.1 ± 8.6	18.1 ± 9.2	24.0 ± 8.8	30.7 ± 5.3	31.1 ± 7.4	38.7 ± 10.2
EARCAL	AirSoft	21.6 ± 11.3	21.6 ± 11.9	22.1 ± 13.9	23.0 ± 12.4	26.3 ± 10.0	31.2 ± 9.6	34.0 ± 13.9
HLI	Airsoft	20.7 ± 9.8	21.0 ± 10.2	22.3 ± 9.7	23.2 ± 8.1	27.3 ± 6.5	31.1 ± 8.3	35.9 ± 11.7
LARI	AirSoft	19.5 ± 9.4	22.1 ± 10.1	24.9 ± 11.3	24.6 ± 10.0	29.4 ± 8.4	36.1 ± 10.0	38.1 ± 11.7
NIOSH	AirSoft	20.3 ± 10.6	20.5 ± 11.1	21.9 ± 11.2	20.4 ± 9.7	27.2 ± 8.3	32.6 ± 10.2	36.0 ± 13.5
USAARL	AirSoft	16.5 ± 12.2	16.3 ± 12.0	17.3 ± 12.0	17.1 ± 11.3	22.5 ± 10.5	23.8 ± 10.9	27.3 ± 14.3

$$\text{Atten}_{f,1} = [19.2, 27.7, 37.7, 35.7, 33.0, 36.0, 41.7],$$

$$\text{Atten}_{f,2} = [19.0, 23.0, 32.3, 37.8, 34.3, 36.3, 37.8],$$

$$\text{Atten}_{f,A,\text{avg}} = [19.1, 25.3, 35.0, 36.8, 33.6, 36.2, 39.8].$$

The second summation of Eq. (1) yielded 73.1 dB. The *A*-weighted attenuation for this subject and device was 107.0−73.1=33.9 dB. The Method-B attenuations were

$$\text{Atten}_{f,B,\text{avg}} = [15.3, 24.1, 31.1, 32.2, 32.8, 30.1, 38.5],$$

which yields an *A*-weighted attenuation of 30.6 dB for this same subject.

The differences in the *A*-weighted attenuations between Methods A and B were determined for each subject to facilitate comparison between methods,

$$\Delta_{\text{A–B}} = A_{\text{A}} - A_{\text{B}}, \tag{2}$$

where A_A and A_B are the attenuations from Eq. (1). For the example data, the difference was 3.3 dB. The differences in *A*-weighted attenuation were analyzed by grouping the subjects within a laboratory and also by pooling all of the subjects together across all laboratories. The averaged differences and the statistical analyses are presented in Table VII.

In Fig. 1 the *A*-weighted attenuations are compared graphically with box-whisker plots. Method-A data are shown on the left of each pair and Method-B data on the right in accordance with standard alphabetical preference; note that the actual sequence of testing was Method B and then Method A. The vertical length of each box represents the interquartile range, which extends from the 25th to the 75th percentiles (i.e., 50% of all data points are contained within the box). The horizontal line inside the box indicates the median value, and the whiskers depict the 10th and 90th

TABLE VII. *A*-weighted attenuations in dB for Methods A and B, attenuation differences (Δ_{A-B}), standard error, student's *t* test, probability associated with *t*, and lower and upper 95% confidence interval boundaries, across all laboratories and by individual laboratory for each protector.

Protector	Lab	*A*-weighted attenuation Method A	*A*-weighted attenuation Method B	Difference Δ_{A-B}	Standard error	Stud. value	Prob. of *t*	Conf. Int. (lower, upper)
TacticalPro	All Labs	29.2	28.1	1.1	0.33	3.350	0.0203	(0.26, 1.95)
TacticalPro	AFRL	28.5	26.7	1.8	0.71	2.612	0.016	(0.39, 3.32)
TacticalPro	EARCAL	29.0	28.7	0.3	1.35	0.227	0.823	(−2.48, 3.09)
TacticalPro	HLI	31.5	29.2	2.3	0.89	2.588	0.016	(0.46, 4.14)
TacticalPro	LARI	29.1	28.3	0.8	0.46	1.925	0.067	(−0.07, 1.82)
TacticalPro	NIOSH	30.8	29.9	0.9	0.64	1.464	0.157	(−0.39, 2.25)
TacticalPro	USAARL	26.5	26.1	0.4	0.74	0.482	0.634	(−1.18, 1.89)
3M 1427	All Labs	27.3	26.5	0.8	0.58	1.080	0.3296	(−0.87, 2.12)
3M 1427	AFRL	24.9	26.4	−1.5	0.86	−1.743	0.095	(−3.26, 0.28)
3M 1427	EARCAL	28.0	27.1	0.9	1.15	0.807	0.428	(−1.45, 3.31)
3M 1427	HLI	31.4	28.6	2.8	1.07	2.668	0.014	(0.64, 5.05)
3M 1427	LARI	26.1	24.9	1.2	0.79	1.434	0.165	(−0.50, 2.76)
3M 1427	NIOSH	27.1	26.9	0.2	1.06	0.150	0.882	(−2.03, 2.35)
3M 1427	USAARL	25.4	25.2	0.2	1.37	0.145	0.886	(−2.64, 3.04)
AirSoft	All Labs	27.8	23.7	4.1	0.64	6.359	0.0014	(2.43, 5.72)
AirSoft	AFRL	26.2	23.6	2.6	1.21	2.142	0.0430	(0.09, 5.10)
AirSoft	EARCAL	28.6	24.4	4.2	1.85	2.289	0.0316	(0.41, 8.04)
AirSoft	HLI	31.7	24.7	7.0	1.57	4.454	0.0002	(3.74, 10.24)
AirSoft	LARI	29.9	26.9	3.0	1.30	2.372	0.0264	(0.39, 5.78)
AirSoft	NIOSH	27.9	23.7	4.2	1.46	2.912	0.0079	(1.23, 7.26)
AirSoft	USAARL	22.4	19.1	3.3	1.85	1.789	0.0867	(−0.52, 7.13)
Classic	All Labs	31.4	23.8	7.6	1.32	5.696	0.0023	(4.12, 10.91)
Classic	AFRL	37.3	26.4	10.9	1.32	8.231	0.0000	(8.13, 13.59)
Classic	EARCAL	31.9	23.5	8.4	1.47	5.764	0.0000	(5.43, 11.50)
Classic	HLI	35.0	25.3	9.7	1.12	8.731	0.0000	(7.46, 12.09)
Classic	LARI	30.7	24.6	6.1	1.09	5.588	0.0000	(3.83, 8.34)
Classic	NIOSH	30.6	22.5	8.1	1.40	5.789	0.0000	(5.21, 11.01)
Classic	USAARL	22.6	20.8	1.8	1.00	1.800	0.0851	(−0.27, 3.87)
JazzBand	All Labs	22.7	19.6	3.1	0.36	8.591	0.0004	(2.15, 3.99)
JazzBand	AFRL	23.7	21.5	2.2	1.06	2.062	0.0506	(−0.01, 4.38)
JazzBand	EARCAL	20.4	16.8	3.6	1.13	3.164	0.0043	(1.24, 5.93)
JazzBand	HLI	25.9	21.5	4.4	1.00	4.463	0.0002	(2.38, 6.50)
JazzBand	LARI	21.0	17.9	3.1	1.43	2.195	0.0385	(0.18, 6.09)
JazzBand	NIOSH	23.4	21.3	2.1	0.78	2.735	0.0118	(0.52, 3.73)
JazzBand	USAARL	21.7	18.8	2.9	1.40	2.130	0.0441	(0.09, 5.86)
dB Blocker	All Labs	29.1	26.7	2.4	0.31	7.554	0.0006	(1.56, 3.18)
dB Blockers	AFRL	31.7	30.0	1.7	1.57	1.083	0.2901	(−1.55, 4.94)
dB Blockers	EARCAL	30.1	26.5	3.6	2.37	1.519	0.1424	(−1.30, 8.51)
dB Blockers	HLI	28.7	26.2	2.5	1.15	2.163	0.0411	(0.11, 4.87)
dB Blockers	LARI	30.3	27.5	2.8	1.43	1.851	0.0771	(−0.31, 5.61)
dB Blockers	NIOSH	28.8	26.4	2.4	2.07	1.133	0.2687	(−1.93, 6.61)
dB Blockers	USAARL	25.1	23.7	1.4	1.85	0.774	0.4469	(−2.39, 5.25)

percentiles. The points depict the individual subject results outside the 10th and 90th percentiles. In panel A of Fig. 1, the Peltor Tactical Pro earmuff yielded essentially identical results for Methods A and B for four laboratories and statistically significant increases in attenuation for Method A over Method B for the AFRL and HLI laboratories. In most cases, the lowest attenuations were eliminated with the training of the naïve subjects. Therefore, the differences can be attributed to the test subjects' lack of experience with hearing protector fitting and testing. The difference for the Peltor product was small yet statistically significant with $\Delta_{A-B} = 1.10$, $p = 0.0203$, CI=(0.26, 1.95), where p is the probability of significance and CI are the upper and lower limits of the 95% confidence interval.

Attenuation values for the 3M 1427 earmuff were higher for Method A, although the differences between the two methods were slight. The HLI laboratory exhibited the greatest difference between Methods A and B in the lower frequencies. The HLI laboratory was the only laboratory where the difference from zero was statistically significant at p

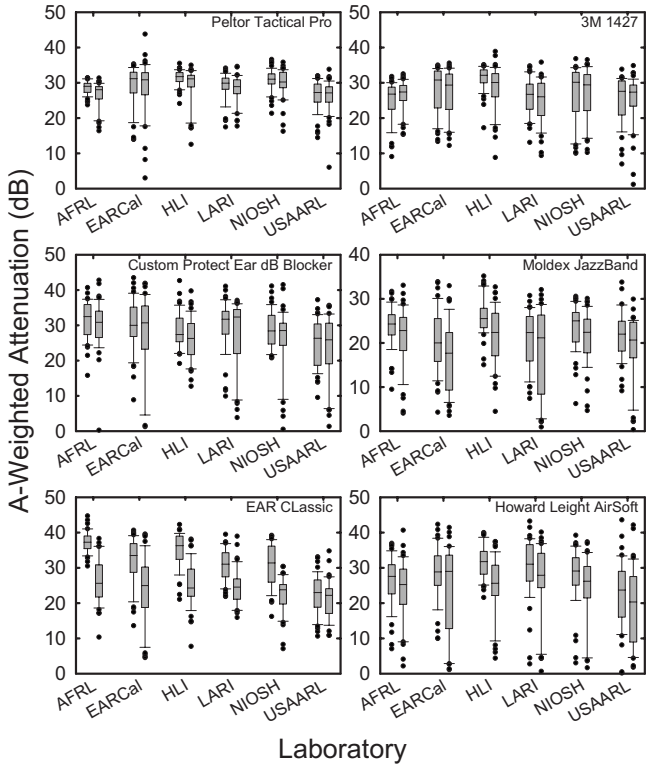

FIG. 1. The *A*-weighted attenuation estimated from the REAT for each method, protector, and laboratory. In each set of paired results, Method A attenuations are displayed on the left and Method B on the right. Each box plot gives the median, 25th and 75th percentiles; the error bars indicate the 10th and 90th percentiles; individual point represent the attenuations observed on individual trials for subjects outside the 10th and 90th percentiles.

<0.05: $\Delta_{A-B}=2.84$, $p=0.014$, CI=(0.64,5.05). Overall, the 3M product did not exhibit a statistically significant difference between Methods A and B.

The Custom Protect Ear dB Blocker results for the overall laboratory and individual laboratory difference analyses exhibited different trends in the improvement of the *A*-weighted attenuation between Methods B and A. Five of the laboratories showed no statistically significant improvement; the confidence interval includes 0. The HLI laboratory exhibited a significant improvement: $\Delta_{A-B}=2.49$, $p=0.04$, CI=(0.11,4.87); the confidence interval excludes 0. The overall laboratory difference analysis exhibited a statistically significant improvement, $\Delta_{A-B}=2.37$, $p=0.0006$, CI =(1.56,3.18). The attenuations at individual frequencies improved between 1 and 5 dB while the standard deviations decreased about 1–2 dB (see Tables V and VI). The change in attenuation based on 24 samples from a single laboratory was not significant; however, pooling the subjects' results (thereby increasing the sample size to 144) causes incremental improvements in attenuation to become significant when exhibited in a larger population.

The JazzBand canal caps exhibited significant improvement between Methods B and A. Only the AFRL subjects did not demonstrate statistically significant improvement at the $p=0.05$ level, but the probability was nearly significant: $\Delta_{A-B}=2.19$, $p=0.0506$, CI=(−0.01,4.38). Δ_{A-B} for the other laboratories ranged between 2.12 dB at NIOSH and 4.44 dB at Howard Leight. Overall, the improvement was $\Delta_{A-B}=3.07$, $p=0.004$, CI=(2.15,3.99). The additional instruction in Method A significantly improved the performance of naïve subjects.

The E·A·R® Classic® earplugs exhibited the greatest effect due to testing under Method A versus Method B. Overall, $\Delta_{A-B}=7.52$, $p=0.0023$, CI=(4.12,10.91). The Classic also exhibited the greatest variability across laboratories. The AFRL laboratory measured *A*-weighted Method-A attenuations above 30 dB for all of its test subjects: $\Delta_{A-B}=10.88$, $p<0.0001$, CI=(8.13,13.59). However, USAARL exhibited no significant improvement in Method-A testing: $\Delta_{A-B}=1.80$, $p=0.851$, CI=(−0.27,3.87). From the working group's discussions of the results, the AFRL laboratory was determined to have scrutinized the subjects under Method A the most carefully prior to commencing Method-A testing. However at the USAARL laboratory, subjects were trained to properly fit the E·A·R® earplugs but the experimenter did not monitor the subjects' fit prior to Method-A testing. Thus, experimenter involvement was an essential element for increasing the attenuation under Method-A protocol. As a result, ANSI S12 Working Group 11 deliberated extensively to improve the procedure to make the requirements more explicit and less open to interpretation.

Finally, the Howard Leight AirSoft earplug showed significant improvement between Methods B and A: $\Delta_{A-B}=4.07$, $p=0.0014$, CI=(2.43,5.72). Like the Classic, only one laboratory did not demonstrate significant improvement. The HLI laboratory had the greatest improvement, $\Delta_{A-B}=6.99$, $p=0.0002$, CI=(3.74,10.24).

C. Statistical analysis

The statistical power calculation developed by Murphy *et al.* (2004) for the previous interlaboratory study was used in the present data analysis. Repeatability is defined as testing the same product under identical testing conditions (i.e., same subjects, equipment, and environment). Reproducibility can be defined in two ways. Between-subject reproducibility is the expected variability if a different group of subjects were to be recruited and tested in the same laboratory with the same equipment and environment. Between-laboratory reproducibility is the expected variability if the product were to be tested in a different laboratory that complies to the same standard testing protocol.

A multi-level analysis of variance (ANOVA) (Netter *et al.*, 1990) was used to estimate the standard deviations for laboratory, subject, and trial effects. The statistical model was

$$Y_{ijk} = \mu + \text{Trial}_{k(ij)} + \text{Subject}_{j(i)} + \text{Lab}_i, \quad (3)$$

where Y_{ijk} is the measured attenuation of a given trial, subject, and laboratory, μ is the overall attenuation, $\text{Trial}_{k(ij)}$ is the random term for the kth trial within the jth subject and ith laboratory, $\text{Subject}_{j(i)}$ is the random term for the jth subject within the ith laboratory, and Lab_i is the random term for the ith laboratory. Variance components and means were estimated with the MIXED procedure in SAS® (SAS, 2007). The method of estimation was residual maximum likelihood. The standard deviations for within-subject repeatability,

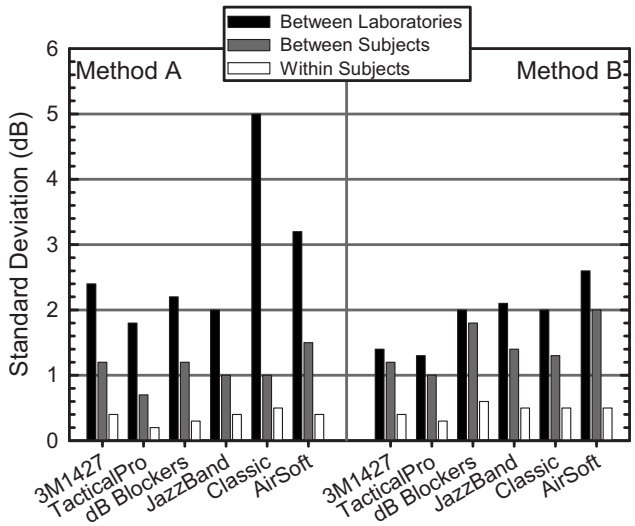

FIG. 2. The within-subject, between-subject, and between-laboratory standard deviations for the *A*-weighted attenuations of each protector using Methods A and B.

between-subject reproducibility, and between-laboratory reproducibility have particular meaning when considering repeated tests of the same product. Between-subject reproducibility is the variability one might expect if the product were tested in the same laboratory with the same subjects and identical conditions. Between-laboratory reproducibility evaluates the testing of a different group of subjects in another laboratory applying the same testing protocol (i.e., similar equipment, psychophysical paradigm, and training).

Within-subject standard deviation was calculated with the equation,

$$\sigma_{\text{within-subject}} = \sqrt{\frac{\sigma^2_{\text{trial}}}{(n_s n_t)}}, \tag{4}$$

where σ^2_{trial} was the trial-to-trial variance, n_s=24 was the number of subjects, and n_t=2 was the number of trials per subject. Between-subject standard deviation was calculated with the equation,

$$\sigma_{\text{between-subject}} = \sqrt{\frac{\sigma^2_{\text{subject}}}{n_s} + \frac{\sigma^2_{\text{trial}}}{(n_s n_t)}}, \tag{5}$$

where $\sigma^2_{\text{subject}}$ was the subject-to-subject variance. Between-laboratory standard deviation was calculated with the equation,

$$\sigma_{\text{between-laboratory}} = \sqrt{\sigma^2_{\text{laboratory}} + \frac{\sigma^2_{\text{subject}}}{n_s} + \frac{\sigma^2_{\text{trial}}}{(n_s n_t)}}, \tag{6}$$

where $\sigma^2_{\text{laboratory}}$ was the laboratory-to-laboratory variance. The standard deviation estimates of $\sigma_{\text{within-subject}}$, $\sigma_{\text{between-subject}}$, and $\sigma_{\text{between-laboratory}}$ were calculated in the multi-level ANOVA, and are presented in Fig. 2. By definition, the standard deviation for any given protector increases progressively as more variance terms are added to the computation. Thus, within-subject standard deviations are the smallest and between-laboratory standard deviations are the largest.

1. Within-subject variability

The within-subject standard deviations were between 0.2 and 0.8 dB for both Methods A and B (see Table VIII). This performance statistic was slightly higher for Method B compared to Method A. The JazzBand and dB Blockers exhibited less variability at all frequencies for Method A. For individual frequencies, the agreement between the Method-A and Method-B within-subject standard deviations were between 0.0 and 0.2 dB. With the exception of the dB Blocker, the *A*-weighted attenuation within-subject standard deviations exhibit either no difference or 0.1-dB differences. The standard deviations of the *A*-weighted attenuations tend to be close to the values in the 1000–4000 Hz region. The improvements of the standard deviations from Method B to Method A were 0.0 dB for the Classic and 3M 1427 devices to 0.1 dB for the AirSoft, JazzBand, and Tactical Pro and 0.3 dB for the dB Blocker. The small decreases for the within-subject standard deviations could be a random effect, a learning effect that results from the subjects having more

TABLE VIII. Within-subject standard deviations in dB for Methods A and B for each protector.

Protector	Method	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	*A*-weight Atten
TacticalPro	A	0.4	0.4	0.3	0.4	0.4	0.3	0.4	0.2
	B	0.4	0.4	0.5	0.3	0.4	0.4	0.4	0.3
3M 1427	A	0.5	0.5	0.5	0.5	0.4	0.4	0.5	0.4
	B	0.5	0.6	0.5	0.5	0.4	0.5	0.4	0.4
dB Blocker	A	0.6	0.5	0.5	0.5	0.4	0.4	0.5	0.3
	B	0.7	0.7	0.7	0.6	0.5	0.6	0.7	0.6
JazzBand	A	0.6	0.5	0.5	0.5	0.4	0.5	0.5	0.4
	B	0.7	0.7	0.6	0.5	0.5	0.6	0.7	0.5
Classic	A	0.7	0.7	0.6	0.5	0.4	0.5	0.5	0.5
	B	0.8	0.8	0.7	0.6	0.5	0.6	0.7	0.5
AirSoft	A	0.6	0.5	0.6	0.6	0.5	0.5	0.6	0.4
	B	0.7	0.6	0.7	0.5	0.5	0.6	0.7	0.5

TABLE IX. Between-subject standard deviations in dB for Methods A and B for each protector.

Protector	Method	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	*A*-weight Atten
TacticalPro	A	0.8	0.9	0.8	0.8	0.7	0.9	0.8	0.7
	B	0.9	1.1	1.1	0.9	0.8	1.1	1.0	1.0
3M 1427	A	1.3	1.3	1.3	1.3	0.7	1.2	1.0	1.2
	B	1.3	1.3	1.3	1.2	0.9	1.3	1.3	1.2
dB Blocker	A	1.7	1.6	1.6	1.3	1.1	1.2	1.5	1.2
	B	2.0	2.1	2.1	1.8	1.7	1.8	2.2	1.8
JazzBand	A	1.4	1.3	1.2	1.1	1.0	1.1	1.5	1.0
	B	1.7	1.6	1.5	1.4	1.5	1.6	2.0	1.4
Classic	A	1.4	1.4	1.5	1.2	0.8	0.7	1.0	1.0
	B	1.4	1.4	1.6	1.3	1.0	1.2	1.5	1.3
AirSoft	A	1.7	1.7	1.9	1.7	1.2	1.7	1.8	1.5
	B	2.1	2.2	2.3	2.1	1.7	1.9	2.6	2.0

experience with the threshold identification task, or an effect of the training in fitting the protector. The study was not designed to test for such effects.

2. Between-subject variability

The variability between subjects is summarized in Table IX. For earmuffs, the standard deviations were 0.7–1.3 dB. For the Peltor TacticalPro, the standard deviations were consistently 0.1–0.3 dB lower for Method A compared to Method B. For frequencies 125–1000 Hz, the between-subject variabilities for the 3M 1427 were essentially the same for both methods, and had slightly less variability in the high frequencies. For the JazzBand and dB Blocker, the Method-B between-subject deviations were 0.3–0.7 dB greater than Method-A deviations. For the Classic at frequencies 125–1000 Hz, the standard deviations were the same or 0.1 dB lower. At higher frequencies (2000–8000 Hz), the Classic's standard deviations for Method A were 0.2–0.5 dB less than the standard deviations observed for Method B. For the AirSoft, the between-subject variability was larger by 0.2–0.8 dB for Method B than for Method A. Differences in standard deviations between Methods A and B less than 1 dB may exhibit statistical significance but have no practical importance.

The standard deviations for the *A*-weighted attenuations tended to correlate with the 1, 2, and 4 kHz estimates and not the 8 kHz deviations. This trend was most evident in the JazzBand, dB Blocker, and AirSoft results. With the exception of the 3M 1427 muff, the *A*-weighted standard deviations were all less for the Method-A than the Method-B testing. For the Classic, the higher frequencies exhibited less variability for Method A than for Method B. The variability approached the lowest for all products at the higher frequencies.

3. Between-laboratory variability

The greatest difference in standard deviations between Methods A and B occurred for the between-laboratory variability (see Table X). Whereas the Method-A standard deviations were smaller than those for Method-B standard deviations for within- and between-subjects, the between-laboratory variability was greater for Method A than for

TABLE X. Between-laboratory standard deviations in dB for Methods A and B for each protector.

Protector	Method	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	*A*-weight Atten
TacticalPro	A	1.8	1.4	2.8	2.4	2.2	2.8	2.4	1.8
	B	2.1	1.6	2.6	2.2	1.9	2.0	2.1	1.5
3M 1427	A	3.1	2.2	2.7	2.3	2.1	3.5	2.6	2.4
	B	2.0	1.4	1.6	1.9	2.5	2.8	2.4	1.3
dB Blocker	A	1.4	1.3	1.8	2.5	2.9	4.1	3.3	2.2
	B	1.7	2.0	1.6	2.6	2.6	3.6	3.4	2.0
JazzBand	A	3.3	3.1	2.2	2.3	2.6	2.9	3.9	2.0
	B	2.7	2.6	2.2	2.7	2.5	2.4	2.2	2.1
Classic	A	4.8	5.1	6.1	6.3	3.3	3.6	4.2	5.0
	B	1.3	1.2	2.0	2.8	2.1	2.9	3.6	2.0
AirSoft	A	3.9	3.7	4.4	3.3	2.5	3.5	3.7	3.2
	B	2.3	2.7	2.9	2.8	2.8	4.0	4.1	2.6

TABLE XI. Calculated numbers of subjects necessary to achieve 6 dB resolution in the *A*-weighted attenuation from between-subject and between-laboratory variances, for Methods A and B. (Results are shown with one decimal place for illustrative purposes.)

Method	Variance	TacticalPro	3M 1427	dB Blocker	JazzBand	Classic	AirSoft
Method A	Between subject	2.5	6.6	7.3	5.2	5.1	11.1
Method B	Between subject	4.3	7.0	15.3	9.3	7.8	18.5
Method A	Between lab	11.7	26.7	19.0	19.0	131.1	52.2
Method B	Between lab	9.2	8.2	16.7	20.2	17.6	27.5

Method B at most frequencies. The Classic, AirSoft, and 1427 devices had greater standard deviations for Method A than Method B at 1000 Hz and below. The Classic had the greatest standard deviations of all the protectors tested. The standard deviations of the *A*-weighted attenuation for these three devices differed across methods by as little as 0.6 dB to as much as 3 dB. In general, the Method-B standard deviations for earplugs increased with frequency, while the standard deviations for earmuffs and canal-caps did not vary appreciably with frequency. For the Classic and AirSoft earplugs, Method-A standard deviations are less at the higher frequencies relative to the lower frequencies. For the dB Blocker, the Method-A standard deviations increased with frequency.

The standard deviations between Methods A and B for the Tactical Pro, JazzBand, and dB Blocker tended to differ by less than 0.5 dB at most frequencies. The standard deviations for the *A*-weighted attenuations were between 0.1 and 0.3 dB for these three devices. The small differences between the standard deviations for the *A*-weighted attenuations suggest that the low frequencies (125–500 Hz) and the highest frequency (8000 Hz) do not contribute significantly to the overall variance for these protectors. The *A*-weighting applied in the computation has de-emphasized the contribution from the lower and highest frequencies.

D. Number of subjects necessary for a desired resolution

The concept of resolution comes from the ability to distinguish between the central tendency of two distributions. In astronomy, the light from two stars can be resolved only if there is sufficient angular separation between the respective images in the telescope. Noise inherent in the image, either due to atmospheric effects or distortions of the equipment, can increase the apparent size of an object. In the analogous case for hearing protectors, the "image" is the distribution of attenuations. The separation of the central modes, the width of the distributions, and the choices for power and confidence level affect the resolution. In essence, the resolution determines the ability to distinguish between the central tendencies of the attenuation distributions in different tests. Because resolution is inextricably linked to the width of the distribution, a wider distribution will require either a greater separation or more subjects to increase the statistical power of the measurement to permit resolving differences between two sets of data. Resolution is dependent on the power and confidence level. If one chooses a low power (0.8) and confidence level (0.84), the resolution will be lower than if a higher power (0.9) and confidence level (0.99) are chosen. The choice of the confidence level and resolution implies that the observed decibel difference in the means of the respective attenuation distributions can be distinguished with 84% or 99% confidence.

In Murphy *et al.* (2004), the number of subjects was estimated using a 6-dB resolution from ANSI S12.6-1997 (R2002) (2002) and the greatest between-subject standard deviation, which typically occurred at 8000 Hz. The frequency with the greatest variance (largest standard deviation) was assumed to dominate the variance of any rating. Since the present analysis examines the repeatability and reproducibility of an *A*-weighted attenuation, the frequencies having the greatest contribution to the *A*-weighted protected and unprotected sound pressure levels (1, 2, and 4 kHz) dominate the variance. This effect is evident in Table XI and Fig. 3 where the numbers of subjects necessary to achieve a 6 dB resolution (i.e., the ability to detect differences between two means of 6 dB) are plotted. From Murphy *et al.* (2004), the required sample size is

$$N = n_s\left(\sqrt{2}(\text{probit}(1-\alpha)+\text{probit}(1-\beta))\frac{\sigma}{R}\right)^2$$
$$= n_s\left(\frac{2.5966\sigma}{R}\right)^2, \quad (7)$$

where n_s is the number of subjects tested, σ is the standard deviation for repeatability or reproducibility as estimated by Eqs. (5) and (6), and R is the desired resolution (6 dB). The

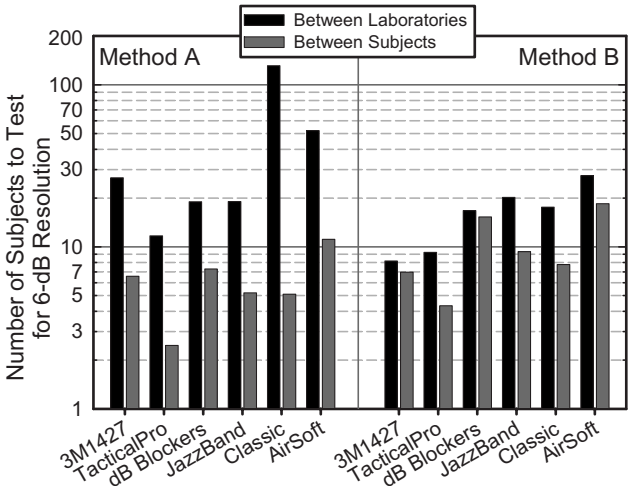

FIG. 3. The power estimates for the number of subjects necessary to achieve a 6 dB resolution for the *A*-weighted attenuations of each protector Methods A and B.

probit function provides the appropriate percentile value from a standard normal distribution for the confidence level, $1-\alpha=0.84$ and power, $1-\beta=0.8$ (see Murphy *et al.* 2004 for further explanation of this approach).

For the Method-A between-subject repeatability (essentially testing the same subject population in the same laboratory), the estimated sample sizes were less than 12 subjects needed to separate the *A*-weighted attenuation results regardless of the protector being tested. The TacticalPro may require fewer than 3 subjects while the AirSoft needs about 11 subjects to achieve a 6-dB resolution.

For the Method-B between-subject repeatability, the AirSoft, JazzBand, and dB Blocker products increased the most. Typically, the estimated sample size was between 10 and 30 subjects if the standard deviations for individual frequencies were used. If the standard deviations based on *A*-weighted attenuations are used, the number of subjects necessary to achieve 6-dB resolution was between 4 and 19 subjects. For the Classic, Tactical Pro, and 3M 1427, 4–8 subjects were required.

The reproducibility between laboratories is considerably poorer than the between-subject repeatability. One should note that the between-laboratory reproducibility includes the between-subject variability and therefore must be greater. From the sample size calculations, Method-B testing would require 9–28 subjects for all products (see Table XI). For the Method-A data, the numbers of subjects required for earmuffs, dB Blocker, and JazzBand were less than 30. However, for the AirSoft and Classic earplugs, 53 and 132 subjects were necessary to achieve a 6-dB resolution, respectively.

The effect of the *A*-weighting can be seen in the power calculation. Note that the between-subject and between-laboratory standard deviations for the *A*-weighted attenuation were more closely correlated with the individual frequency results around 2000 Hz. Particularly the Method-A and Method-B data for the dB Blocker exhibited a minimum at 2000 Hz which correlated closely with the *A*-weighted power estimate. For the AirSoft, the minimum was at 2000 Hz, and the *A*-weighted power was greater than the 2000 Hz value. This suggests that the *A*-weighted power estimate and its associated standard deviation were dominated by those bands (1000, 2000, and 4000 Hz) which have the strongest contribution to the *A*-weighted energy.

IV. DISCUSSION

A. Procedural differences

The testing was completed in August 2006 and the results were initially analyzed in October and November of 2006. ANSI Working Group 11 held three meetings in late 2006 and early months of 2007 to discuss the results and determine modifications to the ANSI S12.6 standard. Laboratory representatives, manufacturers, EPA, and government representatives participated in the deliberations which revealed differences in the implementation of Method-A tests by the six participating laboratories. Particularly, the insertion of the foam earplug was identified as the contributing factor related to different laboratory-to-laboratory experimenter influence under Method A. Requiring subjects to achieve a minimum of 75% insertion is allowed under experimenter-supervised tests. From earlier in-house tests of the E·A·R® Classic®, the experimenter and the laboratory manager at AFRL determined that this depth was necessary to protect personnel in high noise environments (Hall *et al.*, 2005). The fact that AFRL achieved the greatest attenuation of all the laboratories highlights the ability of the product to achieve maximum attenuation with deeper insertion.

The AirSoft also benefited from increased insertion depth. One might be tempted to attribute the highest Method-A attenuations at the Howard Leight Laboratory to its ability to use the product; the effect was not due to subject selection since the same initially naïve subjects were used for both Method-A and Method-B tests. The manufacturer bias was not evident in the test results for the Classic tested at E·A·RCAL. Thus, an appeal to manufacturer bias cannot be substantiated. If the outlier subjects (those who clearly had the least attenuation) for the other laboratories were removed from the distribution, the mean would be increased and standard deviations reduced, yielding better agreement with the results from the Howard Leight Laboratory. Rather, the improved performance at the Howard Leight Laboratory was likely a function of the poorly performing naïve subjects being better motivated and more conscientious of the insertion process during Method-A testing.

B. Fitting of protector types

In an effort to improve the uniformity of results collected with the different protocols, various schemes were investigated to remove outlier subjects. However, the scrutiny of the working group did not yield adequate methods to algorithmically remove outlier data. Instead, the outliers must remain in the data sets and explanations for the aberrant results must be sought.

First, the type of protector must be considered. Earmuffs comprise one category; the custom protectors and semi-aural inserts (i.e., canal-caps), although seemingly different, comprise another. Insert earplugs can be separated into two categories. Essentially, all existing protectors can be fit into these categories regardless of the electronic enhancements or other features of the protector.

Earmuffs must make a seal with the skull around the ear. So long as hair, jewelry, and head shape do not interfere with the ability of the ear cushion to seal against the skull, the protector's attenuation is governed by the volume and mass of the earcup, the compliance of the cushion, the ability to conform to an irregular shape, and the headband clamping force. Differences in these characteristics influence the overall attenuation and possibly have small effects on the variability of the attenuations across subjects. For example, the Peltor Tactical Pro has a slightly more compliant cushion than the 3M 1427 muff. As well, the headbands for the two muffs were worn in different positions (the 3M 1427 was worn behind the head, and the Peltor was worn over the head). However, these characteristics are largely independent of the user, and therefore the differences between the trained and untrained tests were minimal.

The second category, canal-caps and custom products, has a similarity in that once the canal is sealed, little additional attenuation can be achieved (i.e. deeper insertion is not possible). The canal-cap seals the entrance of the ear canal, and the custom protector is designed to extend into the canal. If a custom product is manufactured from a deeper impression, then an increased attenuation will be likely (Hall *et al.*, 2005). Assuming the dB Blockers were manufactured to the same tolerances, the nominal length of the ear canal portion should provide about the same amount of attenuation. For those subjects already achieving good attenuations, the additional experimenter instruction (Method A) did not produce an appreciable improvement compared to the attenuations measured with Method B (see Fig. 1, 10% whisker). For the lower attenuations obtained with Method B, the poorly-sealed ear canals were able to be sealed once the subjects were instructed by the experimenter.

Some banded devices are designed to have the earplug inserted into the ear canal. In this case, attenuation testing results would be expected to resemble those of earplugs rather than canal-caps. If products have a limitation on how far they might be inserted or how the seal with the ear canal is created, then training will improve the poorly-fit protectors but little improvement can be expected for the well-fit protectors.

A third category of protector, premolded earplugs, typically achieves a seal through pressing a single flange or multiple flanges against the canal walls. The seal created by the flanges is the main determiner of the product's attenuation rather than transmission through the body of the earplug. If the flanges fail to adequately seal to the ear canal walls, then the dominant transmission path will be through the leak(s) around the flange. Once the flanges have achieved a patent seal, insertion of the product further into the ear canal will do little to improve or increase attenuation.

The AirSoft product tested in this study is not the same product currently on the market. The product tested in the study had an unbaffled, air-filled bladder within the body of the plug; the current product has been redesigned to incorporate internal baffles. The results may generalize to other similar premolded earplugs. In particular, the previous interlaboratory study examined the V-51R and EP100 earplugs (Royster *et al.*, 1996). The maximum attenuations published by Murphy *et al.* (2004) did not change appreciably between the informed-user and naïve subject-fit tests. Thus, users of premolded plugs in the lower quartile of the study will benefit most from training and instruction.

A fourth category of protector, formable earplugs, creates a seal between the lateral surface of the earplug and the ear canal walls. For the Classic®, the attenuations for both the lowest and highest quartiles improved significantly between Method-A and Method-B tests (see Fig. 1). The improvements for the subjects in the highest quartiles indicate the training was effective and necessary to achieve better attenuations from a foam roll-down earplug. As more of the lateral surface of the plug makes contact with the ear canal wall, the attenuation increases. This finding has been observed in the previous interlaboratory study (Murphy *et al.*, 2004) and explicitly tested in a study conducted by the Air Force (Hall *et al.*, 2005).

During the working group deliberations, the test methods used at AFRL were found to require the subjects insert at least 75% of the Classic into the ear canal. The tester marked the protectors at approximately three-fourths of the length and watched the subject during insertion to ensure that the mark was obscured after insertion. Subjects were required to refit the plug if the mark was visible to the experimenter. While the methods employed by AFRL are permitted under the Method-A protocol, they do not reflect the typical use of the product by a trained or untrained user. The user would need a mirror to examine the insertion depth and a second mirror to inspect whether the mark had been obscured. Furthermore, most users will not ask a co-worker to inspect the fit or insertion depth of the earplug. Although the AFRL method was effective in achieving greater attenuation, it does not reflect typical or practical use.

One key result of the current study was the realization that greater consistency of experimenter instruction of subjects would improve the between-laboratory reproducibility of Method A. The working group has invested substantial time in revising ANSI S12.6 in an attempt to accomplish this, changing Method A from an an experimenter-supervised fit to a trained-subject fit protocol, removing the influence of the experimenter on the fitting of the device during the actual attenuation measurement.

C. Application to regulatory issues

Under the OSHA Hearing Conservation Amendment (OSHA, 1983), workers are enrolled in a hearing conservation program when noise exposures equal or exceed an 8 h time-weighted average sound level of 85 dB(A). Training in the use of hearing protection must be provided, and workers are required to be refitted and retrained whenever they suffer a standard threshold shift and provided with hearing protectors offering greater attenuation if necessary. While the quality of training may vary greatly across companies, those workers that have received it would no longer qualify as naïve subjects. The majority of hearing protection is sold to industrial hearing conservation programs and not to the average consumer (Frost and Sullivan, 2005). This alone suggests that hearing protectors are being used by trained workers and not untrained, uninformed consumers. Method-A data should be more representative of the typical user than Method-B data, yet Berger *et al.* (1998) provides contradictory evidence suggesting that Method B is the more appropriate technique.

In Table XII, several advantages and disadvantages for Methods A and B are presented side-by-side. One driving force is the creation of a reproducible test method. Widely varying product test results with the same subjects are unacceptable. If the EPA were to audit a manufacturer's product, then reproducibility within a laboratory is paramount. This study found better between-subject reproducibility with Method A yet better between-laboratory reproducibility with

TABLE XII. Comparison of advantages and disadvantages between Methods A and B.

Topic	Method A	Method B
Between-subject reproducibility	X	
Between-laboratory reporducibility		X
Testing cost	X	
Testing expediency	X	
Real-world applicability		X
Rank-ordering of result		X
Necessity of derating NRR		X
Explainability of test results	X	
Inherent device performance	X	
Dual number rating	?	?
International applicability	?	?

Method B, though the latter apparent advantage of Method B is helped by the inherently greater between-subject variance that results from using naive subjects.

Testing cost and expediency of testing a product either in an in-house laboratory or an independent laboratory are important considerations for the EPA. The cost and speed of testing will be lower with Method A since the time to recruit, qualify, and maintain a Method-B panel constitutes the initial visit of the subject to the testing laboratory. The six laboratories in this study had varying success in getting subjects to complete the entire series of tests. In some cases, a subject chose to exit the study with only one more Method-A test to complete. A new subject had to be qualified, trained in the task, and run through all of the protector conditions to replace the subject who left. If experienced subjects were permitted, then the subject could have been replaced from the pool of previously qualified subjects.

The EPA should specify a policy regarding the reuse of subjects when testing a product for labeling purposes. Under Method B, the working group has determined that,

> "Once a subject has been accepted in an inexperienced-subject fit evaluation in a given facility, s/he may participate for a lifetime maximum of 30 separate inexperienced-subject fit tests, each test consisting of 2 trials. Of those 30 tests, the total number permissible for earplugs and semi-inserts, or both, shall not exceed 12, and there shall not be more than 4 tests on any one of the following categories: foam, premolded, malleable, semi-insert, and other earplugs, and no more than one test on a custom-molded plug. Subjects shall be excluded from any further inexperienced-subject fit testing of earplugs or semi-inserts once they have viewed video or computer-based fitting instructions during a product test." (ANSI S12.6-2008, 2008).

For Method A, the same subjects may be reused many times for testing products. As subjects develop expertise with the testing paradigm and use of the products under test, they would be expected to provide more consistent test results. With respect to subsequent audit tests or mandated retesting, the reuse of the subject is a topic for debate. If EPA mandates periodic retesting, the lack of any overlap of the testing panel from the initial rating to the retest ensures that the results are statistically independent. Since the period for retesting would likely be more than 2 years, retention of the entire subject panel over that duration is unlikely. An additional argument in favor of nonoverlapping panels would be to increase the numbers of subjects on which the product has been tested. In Murphy *et al.* (2004), the statistical power of the data set increased more by adding subjects than by performing repeated measurements on the same subjects.

With respect to real-world applicability, the Method-A results are expected to be similar to the European SNR [in accordance with ISO 4869-1 (1990) and ISO 4869-2 (1994)] as the subjects would be experienced users of protectors, while workers that regularly use HPDs are undoubtedly experienced, many wear the protectors only to comply with company policy as mandated by federal and state regulations. Those workers achieving inadequate levels of protection (poorly-fit) are at increased risk of developing noise-induced hearing loss. Method-B data have been demonstrated to provide a better correlation with real-world attenuations than the current experimenter-fit NRR or attenuations based on ISO 4869-1 (1990) experimenter-supervised fit data (Gauger and Berger, 2004). Knowledge of how a protector is likely to be worn is useful in predicting rates of hearing loss. In addition to being more applicable to real-world performance, the Method-B data would not need to be derated to assess the likely performance in the real world.

Method-A data are more useful in explaining the performance of a device. While we have not focused on the attenuation by frequency for the various products, the improved low-frequency performance of the Classic® can be better understood when the results are more consistent across subjects. If a manufacturer designs a protector to achieve maximum attenuation for a predominantly low-frequency, high-noise environment, then using an untrained naïve subject is inappropriate. The quality of the seal becomes more critical as the low-frequency attenuation is decreased dramatically by the presence of small leaks in the seal of a protector. The workers in these environments should be specifically trained in the use of the personal protective equipment and ought to be required to demonstrate adequacy of protection through fit-testing of the protectors. Thus the Method-A data should be able to better assess the inherent performance that a HPD is capable of providing.

The last two elements in the table are indeterminate at the present time. ANSI has recently published a method to calculate the effective *A*-weighted sound pressure level when a hearing protector is worn (ANSI S12.68, 2007). In this new standard, the use of two numbers to describe the attenuation that "most users can achieve" and that which "motivated users might be able to achieve" is novel. Essentially the rating describes the variation about the mean, recommending the use of ± 0.84 standard deviations corresponding to the 80th and 20th percentiles. When the working group developed the standard and deliberated the results from this study, the initial intent was to create a two-number rating that was applicable to Method-B data. However, the two-number rating is equally applicable to the Method-A data. Instead of a motivated user, the user may be a trained, expert user. The language associated with the higher and lower ratings is nu-

anced; regardless, one can learn from the spread of the two values, with a smaller spread indicating less variation across users and noise spectra. Thus, the occupational hearing conservationist will have another means to judge the performance of a particular protector.

V. CONCLUSIONS

Whether the Method-A or Method-B data are more similar to international standards and ratings is open for consideration. Currently Brazil, Canada, and Australia/New Zealand have adopted the essential elements of a Method-B procedure from the ANSI S12.6-1997 (2002) standard into occupational hearing conservation standards and directives. Almost identical to Method B, ISO has developed a method using naïve test subjects (ISO 4869-5, 2006). The European Union has adopted the ISO 4869-1 and 4869-2 standards for HPD testing and rating (ISO 4869-1, 1990; ISO 4869-2, 1994). The ISO 4869-1 standard is similar to the Method-A approach. Until tests conducted under Method-A conditions are available, the similarity of attenuations with those of ISO 4869-1 is unknown. Since both standards utilize trained subjects, the attenuations should be similar.

The U.S. EPA is anticipated to propose a revision to the hearing protector labeling regulation, 40 CFR 211 Subpart B. The Method-A test protocol is expected to be the proposed method for assessing the performance of passive hearing protectors. At the heart of the Noise Control Act is the need *to provide accurate and understandable information to product purchasers and users regarding the acoustic properties of designated products so that meaningful comparisons with respect to noise emission or noise reduction can be made as a part of a product purchase or use decision* (EPA, 1972). Method B can more nearly represent the anticipated protection of uninformed users. Once a user has experienced the higher attenuation and tactile sensation of a well-fit earplug and has learned techniques to check the fit, the assumption of performance typified by naïve users is potentially erroneous. Furthermore, the question remains as to when an uninformed or naïve user becomes an experienced user with respect to repeated fittings of earplugs. While the attenuation achieved by uninformed, naïve users is important for estimating the risk of noise-induced hearing loss of a group of persons exposed to a given noise, the purpose of the label is to provide the manufacturer a means to inform the user of the acoustic properties of the product. Underestimating the NRR by applying a Method-B protocol could present a disservice to the public and could unfairly disadvantage manufacturers who create high attenuation protectors.

This research, the participating laboratories, and the ANSI S12 Working Group 11 have provided invaluable assistance to the EPA in the preparation of a revised regulation. The original promulgation of 40 CFR 211 Subpart B occurred in 1979. Changes to the fundamental protocol for rating hearing protector performance are now being contemplated in a notice of proposed rule making by the EPA. Future changes may not occur again for several decades. Thus the effort to quantify the differences between ANSI S12.6 Method A and Method B has led to revisions of the test standard that will affect the policy of the United States for years to come.

ACKNOWLEDGMENTS

Portions of this work were supported by the U.S. EPA interagency Agreement No. 75921973-01-0. Each participating laboratory funded the subject payments from their respective internal budgets. Each laboratory had employees without whose assistance and dedication this study would not have been able to be completed: Paul Schley, AFRL; Ron Keiper, E·A·RCAL; Jesse Duran, Howard Leight; Racheal de Favacorte, LARI; Pamela Graydon, Dayna Richards and Benson Davis, NIOSH; Adrian Houtsma, Melinda Hill and Kimberly Vasquez, USAARL. NIOSH also wishes to note the contributions of Christa Themann and Edward Zechmann for their thorough review of the manuscript. As well, the authors wish to thank Lee Hager, Torben Poulsen, and Rickie Davis for their thoughtful comments during the NIOSH peer-review process.

DISCLAIMER

The findings and conclusions in this report are those of the authors and do not represent any official policy of the Centers for Disease Control and Prevention, The National Institute for Occupational Safety and Health, the Environmental Protection Agency, the U.S. Air Force or the U.S. Army. Mention of company names and products does not constitute endorsement by CDC, NIOSH, EPA, U.S. Air Force or the U.S. Army.

ANSI S12.6-1997 (R2002) (**2002**). American National Standard for the Measuring Real-Ear Attenuation of Hearing Protectors, American National Standards Institute, New York.

ANSI S12.68 (**2007**). American National Standard Methods of Estimating Effective *A*-Weighted Sound Pressure Levels When Hearing Protectors, are Worn, American National Standards Institute, New York.

ANSI S12.6-2008 (**2008**). American National Standard for the Measuring Real-Ear Attenuation of Hearing Protectors, American National Standards Institute, New York.

Berger, E. H., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (**1998**). "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. **102**, 665–672.

EPA (**1972**). P. L. 92-574 Noise Control Act of 1972, 86 Stat. 1234, U.S. Environmental Protection Agency, October 27.

EPA (**1978**). CFR 40 211B Hearing Protective Devices. U.S. Environmental Protection Agency, September 29.

Frost and Sullivan (**2005**). "U.S. markets for industrial hearing protection products," Technical Report, December 13 (Frost and Sullivan, San Antonio).

Gauger, D., and Berger, E. H. (**2004**). "A new hearing protector rating: The noise reduction statistic for use with *A*-weighting (NRS_A)," Technical Report No. E-A-R 04-01/HP, American National Standards Institute, New York. This report can be found online at the following URL web URL http://www.e-a-r.com/pdf/hearingcons/TO4_01EPA.pdf.

Hall, J., Mobley, F., McKinley, R., and Schley, P. (**2005**). "New directions for custom earplugs," in 2005 Conference and Exposition on Noise Control Engineering Rio de Janeiro, Brazil, p. 89.

ISO 4869-1 (**1990**). "Acoustics—Hearing Protectors Part 1: Subjective method for the measurement of sound attenuation," International Organization for Standardization, Geneva.

ISO 4869-2 (**1994**). "Acoustics—Hearing Protectors Part 2: Estimation of

effective *A*-weighted sound pressure levels when hearing protectors are worn," International Organization for Standardization, Geneva.
ISO 4869-5 (**2006**). "Acoustics—Hearing Protectors Part 5: Method for estimation of noise reduction using fitting by inexperienced test subjects," International Organization for Standardization, Geneva.
Johnson, D. L., and Nixon, C. W. (**1974**). "Simplified methods for estimating hearing protector performance," Sound Vib. **8**, 20–27.
Kroes, P., Fleming, R., and Lempert, B. (**1975**). "List of personal hearing protectors and attenuation data," Technical Report Publication No. 76-120, U.S. Department of Health Education and Welfare, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati.
Murphy, W. J., Franks, J. R., and Krieg, E. F. (**2002**). "Hearing protector attenuation: Models of attenuation distributions," J. Acoust. Soc. Am. **111**, 2109–2116.
Murphy, W. J., Franks, J. R., Berger, E. H., Behar, A., Casali, J. G., Dixon Ernst, C., Krieg, E. F., Mozo, B. T., Ohlin, D. W., Royster, J. D., Royster, L. H., Simon, S. D., and Stephenson, C. (**2004**). "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**, 311–323.
Netter, J., Wasserman, W., and Kutner, M. H. (**1990**). *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, 3rd ed., Richard D. Irwin Inc., Boston, pp. 970–1001.
OSHA (**1983**). CPL 2-2.35A-29 CFR 1910.95(b)(1) Guidelines for noise enforcement: Appendix A, U.S. Department of Labor, Occupational Safety and Health Administration, December 19.
Poulsen, T., and Hagerman, B. (**2004**). "A Nordic round robin test on hearing protectors. The influence of the sound field on measured REAT attenuation," Acta Acust. **90**, 838–846.
Royster, J. D., Berger, E. H., Merry, C. J., Nixon, C. W., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Mozo, B. T., Ohlin, D., and Royster, L. H. (**1996**). "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.
SAS (**2007**). SAS/STAT Software, 9.0.1 ed., SAS Institute Inc., Carey, NC.