

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

Julia Doswell Royster

Environmental Noise Consultants, Inc., P.O. Box 30698, Raleigh, North Carolina 27622

Elliott H. Berger

E-A-R/Cabot Safety Corporation, Indianapolis, Indiana 46268

Carol J. Merry

National Institute for Occupational Safety and Health, Cincinnati, Ohio 45226

Charles W. Nixon

Armstrong Lab, Wright-Patterson Air Force Base, Ohio 45433

John R. Franks

National Institute for Occupational Safety and Health, Cincinnati, Ohio 45226

Alberto Behar

Behar Noise Control, Scarborough, Ontario M1M 2X8, Canada

John G. Casali

Virginia Tech, Blacksburg, Virginia 24061

Christine Dixon-Ernst

ALCOA, Pittsburgh, Pennsylvania 15219

Ronald W. Kieper

E-A-R/Cabot Safety Corporation, Indianapolis, Indiana 46268

Ben T. Mozo

U.S. Army Aeromedical Research Laboratory, Fort Rucker, Alabama 36330

Doug Ohlin

U.S. Army Center for Health Promotion and Preventive Medicine, Aberdeen Proving Ground, Maryland 21010

Larry H. Royster

North Carolina State University, Raleigh, North Carolina 27695

(Received 7 September 1995; accepted for publication 13 December 1995)

This paper describes research conducted by Working Group 11 of Accredited Standards Committee S12, Noise, to develop procedures to estimate the field performance of hearing protection devices (HPDs). Current standardized test methods overestimate the attenuation achieved by workers in everyday use on the job. The goal was to approximate the amount of attenuation that can be achieved by noise-exposed populations in well-managed real-world hearing conservation programs, while maintaining acceptable interlaboratory measurement variability. S12/WG11 designed two new laboratory-based protocols for measuring real-ear attenuation at threshold, with explicit procedures for subject selection, training, supervision, and HPD fitting. After pilot-testing, S12/WG11 conducted a full-scale study of three types of earplugs and one earmuff tested by four independent laboratories using both protocols. The protocol designated as "subject-fit" assessed the attenuation achieved by subjects who were experienced in threshold audiometry, but naive with respect to the use of hearing protection, when they fit HPDs by following manufacturers' instructions without any experimenter assistance. The attenuation results from the subject-fit method corresponded more closely to real-world data than results from the other protocol tested, which allowed the experimenter to coach subjects in HPD use. Comparisons of interlaboratory measurement variability for the subject-fit procedure to previous interlaboratory studies using other protocols indicated that the measurements with the new procedure are at least as reproducible as those obtained with existing standardized methods. Therefore, the subject-fit protocol was selected for consideration for use in

CONTENTS

Introduction.....	00	1. SF condition.....	00
I. Overview of the problem.....	00	2. IUF condition.....	00
A. Need for a new test method.....	00	F. Optional test of muffs with glasses.....	00
B. Alternative approaches.....	00	G. Anthropometric measurements.....	00
C. Current HPD test procedures.....	00	III. Results.....	00
1. ANSI S3.19-1974.....	00	A. Tabular results.....	00
2. ANSI S12.6-1984.....	00	B. Data analysis.....	00
3. EPA fitting.....	00	C. Interlaboratory variability by fitting instruction	
4. ISO 4869-1:1990.....	00	condition.....	00
D. Measurement reliability (reproducibility).....	00	D. Effect of safety glasses on earmuff attenuation	
E. Testing potential new procedures.....	00	and interlab variability.....	00
1. Fitting techniques.....	00	E. Earmuff band force measurements.....	00
2. Naive hearing protection users.....	00	IV. Comparison of S12/WG11 interlaboratory	
3. Other subject-selection criteria.....	00	variability to that of other studies.....	00
4. Explicit instructions.....	00	A. E-A-R plug data.....	00
5. Test of the HPD-plus-its-instructions.....	00	B. V-51R earplug data.....	00
F. Pilot interlaboratory study.....	00	C. Earmuff data.....	00
G. Full-scale interlaboratory study.....	00	V. Discussion.....	00
II. Method.....	00	A. Suitability of the subject fit protocol for	
A. Experimental design.....	00	standardization.....	00
B. Subjects.....	00	B. Experience with the subject-fit protocol.....	00
C. Hearing protectors.....	00	C. HPD performance as a function of fitting	
D. Apparatus.....	00	instruction.....	00
E. Fitting instruction conditions.....	00	VI. Conclusions.....	00
		VII. Next steps for S12/WG11.....	00
		Acknowledgments.....	00

INTRODUCTION

One of the most vexing problems in hearing conservation is the discrepancy between the labeled attenuation of hearing protection devices (HPDs) and the much lower attenuation actually attained by workers using these devices on the job. At present, manufacturers of HPDs sold in the United States are required by the U.S. Environmental Protection Agency (EPA) to test their products according to a laboratory protocol to obtain a single number rating of hearing protector effectiveness—the Noise Reduction Rating or NRR (EPA, 1979)—which is printed on the package label. Consumers of hearing protection depend upon the NRR when judging the purported relative effectiveness of various devices against noise. Reviews of numerous studies have clearly shown that the experimenter-fit laboratory procedure (ANSI, 1974) currently used for determining the attenuation of hearing protectors results in high NRRs for these devices that greatly overestimate the HPD noise reduction obtained by workers on the job (Berger, 1993a; Berger *et al.*, 1996). Even the Occupational Safety and Health Administration (OSHA) recognizes the problem, as evidenced by a derating scheme which compliance officers apply in certain circumstances (OSHA, 1983). Accurate values for estimating HPDs' real-world attenuation are needed by occupational health professionals and others involved with hearing conser-

vation for noise-exposed personnel. Reliance on inflated NRRs may result in underprotection of workers and needless hearing loss, increased employer liability for Workers' Compensation claims and tort suits, as well as decreased use of engineering noise controls to reduce worker noise exposures (Dixon-Ernst and Behar, 1993).

I. OVERVIEW OF THE PROBLEM

Accredited Standards Committee S12, Noise, assigned its Working Group 11, "Field Effectiveness and Physical Characteristics of Hearing Protectors," to develop a testing procedure that would yield more realistic predictions of the noise reduction that properly trained and motivated workers may receive from using their hearing protectors within an effective hearing conservation program. This paper describes research conducted by the members of S12's Working Group 11 (designated S12/WG11) to define a laboratory-based method for estimating real-world attenuation.

A. Need for a new test method

Available research has clearly and consistently demonstrated a substantial divergence between laboratory and real-world data. Berger (1993a) summarized 20 published stud-

ies, and more recently Berger *et al.* (1996) included two additional studies in the database, which now spans seven countries and over 90 industries with a subject count of approximately 2900. Overall, field attenuation values for earplugs average about 25% of the published U.S. laboratory values (range is from 5% to 52%); earmuff attenuation averages about 60% of the laboratory values (range is 47% to 76%). This divergence is especially troubling considering the importance that many purchasers of HPDs place upon published attenuation data and their confidence in the data's validity.

One might hope to resolve laboratory/real-world discrepancies by increasing field performance through improved occupational hearing conservation practices in the areas of fitting and training of HPD users, enforcement of correct HPD utilization, proper HPD care and regular replacement of worn devices, education and motivation of the work force, and program management. However, experimenter-fit methods yield overly idealistic attenuation data that can never be matched by groups of real-world HPD wearers. Exceptional, highly motivated *individual* wearers can achieve protection consistent with experimenter-supervised-fit data. However, it has never been demonstrated that any *group* of workers operating under real-world constraints can attain and maintain attenuation matching experimenter-supervised-fit values. This attenuation is unlikely ever to be achieved by groups of real workers, even if more comfortable and easier-to-use devices were available and hearing conservation program practices were improved. Therefore, HPD testing methods must be devised to yield data which more closely correlate with existing or potential performance for groups of real-world wearers.

In 1987 the existing S12/WG26, Hearing Protector Characteristics, was tasked with this new assignment: "(to) explore the problems inherent in using optimum-laboratory real-ear attenuation data to estimate achievable and/or typical workplace protection, (and) propose a plan of action to identify or develop laboratory and/or field procedure(s) that yield useful estimates of field performance." In conjunction with this broadened scope, the name of the working group (WG) was changed to "Field Effectiveness and Physical Characteristics of Hearing Protectors." In 1990 the WG was redesignated as S12/WG11. The WG's new mandate arose from the growing awareness and concern of the ASA Standards Director and the chair of S12 regarding the poor correspondence between NRRs and field performance, and the need for buyers and users of hearing protectors to have a reliable and accurate means of predicting the amounts of protection they might achieve in practice.

B. Alternative approaches

In deciding how to estimate real-world performance, S12/WG11 considered two approaches: *predictive* methods using laboratory measurements, and *monitoring* methods using field measurements (as in existing real-world studies). Although the WG plans to define a standard monitoring method as a future project, members agreed to develop a predictive method first, because correcting the potentially

misleading labeled attenuation values in today's marketplace was felt to be the higher priority. The predictive methods suggested were of two types:

- (1) Utilize ANSI S12.6 electroacoustic and psychoacoustic procedures, but modify the subject selection, supervision, training, and fitting procedures and the required sample size and number of trials to model more closely the field application of HPDs, so that results would provide a more valid estimator of real-world attenuation, or
- (2) devise a derating scheme to be applied to existing ANSI S12.6-1984 data.

Method 1, prediction, was selected in favor of Method 2, derating, because there were insufficient data available to provide HPD-specific derating estimates, and because there is also reason to believe that the rank ordering provided by current real-ear attenuation at threshold (REAT) standards does not correspond to relative real-world performance (Berger *et al.*, 1996). Thus it was questionable whether any derating scheme could be devised short of remeasuring all HPDs, a procedure no less time consuming than implementing a new predictive test method. Recent research results have indicated that it is possible to model field usage under laboratory conditions (Berger, 1988; Casali and Epps, 1986; Casali and Park, 1991). For example, Berger has demonstrated that subject-fit data for large groups ($N > 50$) of primarily naive HPD wearers provided a much closer correspondence to typical, or even optimum, real-world data than do standard laboratory tests with select, well-trained, closely supervised, and highly motivated test subjects.

In developing procedures to estimate real-world attenuation, the WG had to select the type of field data to be predicted. In practice, a wide range of HPD attenuation values may be observed, from essentially no attenuation at all for devices poorly fitted by untrained and unmotivated users, to much higher levels of protection that may be attained in the most successful HCPs under ideal conditions. The WG decided to try to estimate "achievable" results in a well-run HCP, defined as the attenuation obtained in the top 10%–20% of the available real-world studies, or the attenuation that should be obtained by an informed and motivated work force in an effective HCP. The validity of the estimates could be assessed by comparing laboratory-measured values arrived at using the WG's proposed protocol, to values derived by examining the 22 real-world studies previously analyzed by Berger *et al.* (1996).

C. Current HPD test procedures

The HPD attenuation testing method that is most commonly used and standardized worldwide is that of real-ear attenuation at threshold (REAT) using human subjects. It is generally accepted that REAT data accurately represent the attenuation achieved by the subjects who were tested, for the way in which they wore the devices under test (Berger, 1986). How test subjects fit and wear the devices is a critical determinant of the attenuation measured, especially for earplugs. These factors are influenced by the experimenter who selects, trains, supervises, motivates, and fits the subjects—in short by the experimenter's own preconceptions about how

the HPD should be used and is expected to perform. Different standards allow varying degrees of experimenter influence.

1. ANSI S3.19-1974

In the 1974 version of the ANSI REAT standard, two types of fitting were specified, subject fit and experimenter fit. Subject fit was denoted an “average” fit and experimenter fit as a “best” fit. Subject-fit testing called for the experimenter to instruct the subject according to the manufacturer’s instructions and to select the proper size HPD. Fitting noise was presented and the listener was directed to fit the device for minimum perceived noise.

The experimenter-fit version was actually an augmentation of the subject fit, in that once the subject fit was accomplished the experimenter was then told to “personally check each installation to assure a good fit and acoustic seal.” If the experimenter deemed necessary, she or he reinserted or readjusted the HPDs for a best fit.

The subject-fit procedure of the WG’s current protocol differs from the 1974 standard in certain aspects, and also in that specific details are more explicit and therefore less open to interpretation (see Sec. II E). Instead of being instructed to adjust the HPD for maximum noise exclusion, subjects are told to fit the HPD to the best of their ability in a manner in which they would be likely to use it in a noisy environment. The experimenter is not allowed to provide any assistance whatsoever except to supply to the subject any manufacturer’s written instructions which would normally accompany the product. Experimenters cannot amplify or augment unclear or missing aspects of manufacturers’ instructions, and time limits are placed on how long a subject may practice. No fitting noise is allowed since this would require instructing subjects how to use the noise when fitting the HPD and virtually no manufacturers’ instructions provide such guidance.

2. ANSI S12.6-1984

In the revisions to S3.19, the subject- and experimenter-fit options were eliminated and a new fitting procedure, termed experimenter-supervised fit, was devised. This protocol was similar to the prior S3.19-experimenter-fit procedure except that the refitting, if needed, would be accomplished by the subject and not the experimenter. The goal of the procedure was to “yield optimum performance values, which may not usually be obtained under field conditions.”

3. EPA fitting

Although the nominal requirements of the EPA labeling regulations were to implement the experimenter-fit procedure of ANSI S3.19-1974, the EPA’s specific implementation neglected the required first step of the S3.19-specified experimenter fit (namely, a subject fit of the device), thereby turning the fitting into a pure experimenter-controlled procedure. Thus, according to the current EPA interpretation and test lab practices, an experimenter-fit procedure utilizes the subject as though she or he were a test fixture to which the experimenter applies the HPD being tested.

4. ISO 4869-1:1990

European and ISO standards typically have recognized the importance of comfort and the need for more realistic and real-world achievable fitting conditions. A representative standard is the current version of ISO 4869. It speaks specifically to comfort by calling for subjects to fit the HPD and adjust it for “best attenuation consistent with reasonable comfort.” Another major feature of ISO 4869 is that it gives explicit indication that, once all training of the subject in fitting the device is complete, the HPD is removed; the subject must refit the device on his or her own for the definitive test.

Regardless of the differences between the ISO and ANSI standards, and the fact that European laboratories implementing such procedures generally have found lower mean attenuation values and higher standard deviations than do their U.S. counterparts, the mean attenuation values measured according to ISO still generally are higher than typical results observed in actual occupational settings (Berger, 1988; Berger *et al.*, 1996).

D. Measurement reliability (reproducibility)

In developing a new predictive procedure, S12/WG11 focused on the accuracy or validity of the resultant data; by design, the approach was intended to produce lower, more realistic, mean values of attenuation than current methods. However, there was concern that the new procedure might, as a consequence, result in unreasonably high(er) standard deviations. Low standard deviations may be considered desirable because they suggest good reproducibility by the same HPD users on repeated measurements. To the extent that high standard deviations are indicative of poor reproducibility either by different groups of users on repeated measurements within the same facility, or by different groups of users in different facilities, such standard deviations would be undesirable. But, this is not necessarily the correct inference to draw from high within-test variability on one group of test subjects.

For example, in a comparison of the within-laboratory reproducibility for multiple REAT tests on one brand of foam earplugs in four different laboratories, over an eight-year period of time, Berger (1988) found that the laboratory with the smallest average within-test variability had the worst reproducibility from test to test. The laboratory with the largest average within-test variability had among the best test-to-test reproducibility.

A goal of the new S12/WG11 procedure was to reduce variability between tests and facilities by considering additional factors influencing the spread of data, such as the skill of the experimenter (see Sec. I E 4) and the selection of test subjects. Howell and Martin (1973) demonstrated the importance of between-subject as opposed to within-subject variability, especially for earplugs; standard deviation values will be justifiably high, regardless of test method, for a HPD which does not consistently yield a satisfactory acoustic seal for a variety of different wearers.

Measurement reliability (rather than validity) has been the focus of existing test methods. In spite of this focus, and

even though a standard procedure supposedly was being followed, past studies of interlaboratory variability have indicated that surprisingly large differences can occur in both the means and standard deviations of attenuation (see Sec. IV). Since S12/WG11's mandate was to estimate attainable real-world attenuation, the decision was made to consider *both* validity and reliability, but to assign principal importance in the evaluation process to the validity or accuracy of the resultant data. The secondary goal was that the interlaboratory variability of the new procedure should be no greater than that found in prior studies.

E. Testing potential new procedures

The WG devised two protocols that incorporated the basic electroacoustic and psychoacoustic procedures of ANSI S12.6-1984 (ANSI, 1984), except that the test frequencies of 3150 and 6300 Hz were eliminated to increase testing efficiency (Berger and Rowland, 1989). Additionally, aspects of the protocols concerning subject selection, training, and supervision were changed to model typical use conditions more closely. To provide greater diversity of subjects, each measurement condition included 24 subjects with two trials per device (compared to ten subjects with three trials per device as in current U.S. test standards).

1. Fitting techniques

The two protocols selected for comparison were designated "subject fit" (SF) and "informed user fit" (IUF). In essence, SF consists of handing the subjects HPDs and the manufacturer's instructions, and asking them to fit devices to the best of their ability in a way in which they would be likely to wear them in a noisy environment. Experimenter involvement is kept to an absolute minimum. This protocol was hypothesized to yield mean attenuation data close to the target of achievable real-world performance, but WG members anticipated that interlaboratory variability might be excessive using SF. The IUF protocol called for the experimenter to interact with the subject (without touching the subject or the subject's HPDs) to demonstrate correct use according to the manufacturers' instructions before the subject finally put on the HPD for testing. Although the WG members hypothesized that IUF would yield higher attenuation values than desired, it was expected that interlaboratory variability would be lower in IUF than in SF.

2. Naive hearing protection users

Across labs, the strategy selected to standardize the degree of prior experience subjects were permitted in fitting and using HPDs was to require explicitly defined naive users with no prior HPD test experience or training in the use of HPDs. Listeners with previous audiometric training were not excluded. In fact, the protocol was designed to ensure that all listeners became audiometrically proficient, with demonstration of repeatable open-ear thresholds, prior to hearing protector testing.

3. Other subject-selection criteria

With the small number of subjects that typically can be tested in any one REAT evaluation, it was decided that it would not be feasible to require any specific distributions based upon subject age, earcanal size, anatomical features, or race. However, a gender balance of 50/50% \pm 10% was specified. Subjects were required to meet hearing sensitivity and threshold variability specifications, to be free from visually obvious aural abnormalities, and to be capable of reading the instructions they would be provided. No other selection or rejection criteria were allowed.

4. Explicit instructions

The WG felt that one of the most likely sources of interlaboratory variability was the experimenters themselves—how their experience and judgement influenced the guidance they gave to subjects, and the control they exerted over them. To limit this influence, explicit instructions for working with, training, and fitting the subjects were written into the protocol, including a number of statements that experimenters were required to read verbatim to the subjects.

5. Test of the HPD-plus-its-instructions

Since the goal in modeling real-world performance is to estimate what can be achieved in practice, the HPD should be assessed together with the manufacturer's instructions for its use. The "product" being tested consists of the HPD and any instructions which accompany it; therefore, the protocol limited instructions to only those items explicitly stated by the manufacturer. For example, when testing earplugs, if the instructions failed to indicate that the pinna should be pulled during insertion, or neglected to mention the exact positioning or rotation of the earplug within the earcanal, the experimenter was prohibited from clarifying or expanding the instructions even though his or her own experience might dictate otherwise.

F. Pilot interlaboratory study

In 1990 the WG pilot-tested the draft protocol on two types of earplugs using ten subjects per laboratory in four facilities: E•A•RCALSM Laboratory, Cabot Safety Corp., Indianapolis, IN; NIOSH Taft Laboratories, Cincinnati, OH; U.S. Army Aeromedical Research Lab (USAARL), Fort Rucker, AL; and Auditory Systems Laboratory, Virginia Tech, Blacksburg, VA. The outcome of the pilot study was presented at the 120th meeting of the Acoustical Society of America (Berger *et al.*, 1990), and results from NIOSH Taft Labs were later published (Merry *et al.*, 1992). The results demonstrated that SF mean attenuation values were comparable to available real-world studies, and that subjects achieved significantly greater attenuation in the IUF condition. However, the interlaboratory reproducibility of the mean attenuation data was similar in both SF and IUF conditions. The indications were that the SF protocol yielded the better estimate of real-world attenuation, while still providing acceptable reproducibility. Experience from the pilot study suggested modest refinements in the draft protocol, which was subsequently revised (Berger, 1990).

G. Full-scale interlaboratory study

A full-scale study, in which three earplugs and one earmuff were tested in four laboratories, was begun in late 1990 and concluded in December 1991. Earplugs were emphasized over earmuffs because they exhibit greater change in achieved attenuation when subject selection, training, fitting, and supervision procedures are varied. Three of the four facilities involved in the pilot study participated in the full-scale study; Virginia Tech was unable to take part and was replaced by Armstrong Laboratory, Wright-Patterson Air Force Base, OH (WPAFB). Various aspects of the full-scale study were discussed in a special session at the 126th Meeting of the Acoustical Society of America (Berger, 1993b; Dixon-Ernst and Behar, 1993; Franks and Casali, 1993; Franks and Merry, 1993; Nixon, 1993; Ohlin, 1993; Royster and Royster, 1993).

This paper presents the attenuation results and interlaboratory variability of the full-scale study. Separate papers in preparation by WG members will address the validity with which the SF data estimate representative real-world attenuation, as well as the most efficient numbers of subjects and trials per device needed to achieve reliable results (Franks *et al.*, 1996a and 1996b, in preparation).

II. METHOD

A. Experimental design

Within each of four laboratories, 24 subjects participated in a repeated-measures design to test attenuation for each of four HPDs in each of two fitting instruction conditions: subject fit (SF) and informed user fit (IUF), described fully below. After an initial session for hearing screening and audiometric training, each subject made eight visits to the lab, as shown in Table I. In each visit, two different HPDs were tested in separate sessions, each including two trials pairing sets of unoccluded and occluded thresholds (sound-field thresholds without and with HPDs). Subjects received a brief rest break between the two sessions of a visit. Visits were separated by a minimum of 6 h, and all eight visits were required to take place within 21 days.

The order in which HPDs were tested was counterbalanced across subjects, yielding 24 unique sequences. Within each lab, six subjects tested each HPD first, second, third, and fourth. For each individual subject, the same sequence of HPDs was repeated in each experimental condition, with the order of unoccluded and occluded threshold measurements counterbalanced. The repetition of testing *within* each fitting instruction condition (SF 2 and IUF 2) provided information to test the effect of additional practice in using HPDs without any additional instruction.

B. Subjects

Each laboratory recruited 24 subjects who were naive with respect to use of HPDs. Naive subjects were defined as those who had never participated in measurements of HPD attenuation, had never received personal instruction in the use of HPDs, had never worn HPDs on the job, and had not worn HPDs or any other type of earplug (such as swimming

TABLE I. Repeated-measures experimental design. Within each lab, each subject was assigned a unique sequence of HPDs for testing.

FITTING INSTRUCTION CONDITION	VISIT	HPD	TRIAL	
SF1	1	1	1	
		2	2	
		3	1	
		4	2	
	2	1	1	
		2	2	
		3	1	
		4	2	
SF2	3	(repeat of visit 1)		
	4	(repeat of visit 2)		
	IUF1	5	1	1
			2	2
6		3	1	
		4	2	
IUF2	7	(repeat of visit 5)		
	8	(repeat of visit 6)		

plugs) more than twice in the past month or more than five times in the past year. In addition, subjects were given an air-conduction audiogram to ensure that their pure-tone hearing thresholds were no greater than 25 dB from 125 to 8000 Hz. Their ears, as determined by otoscopic inspection, had to be free of impacted cerumen or obvious signs of infection. Subjects also were required to demonstrate a level of literacy and visual acuity sufficient to be able to read material similar to the instructions packaged with the hearing protectors. Within each lab, the gender balance of subjects had to be between 50/50% and 60/40%. In the initial training session, subjects performed a minimum of five unoccluded sound-field audiograms until the thresholds measured on the last three audiograms were consistent within a range of 6 dB at all seven frequencies tested. If necessary, additional training audiograms were permitted until the subject either met the requirement for threshold reliability in three consecutive audiograms or was rejected.

In order to obtain data more representative of the real-world working population, no other subject-selection criteria were permitted. For example, the ANSI S12.6-1984 protocol permits an experimenter to reject a subject for whom a good fit is difficult to obtain because of some anatomical extreme such as large pinnae or unusual earcanals. Such rejections were prohibited in this study. Only obvious anatomical abnormalities that would preclude reasonable use of plugs or muffs were grounds for rejecting a subject.

The final subject pool consisted of 96 subjects, 53 females and 43 males, ranging in age from 16 to 61 years old, with a mean age of 25 years. Several potential subjects were rejected for failing to meet the requirement for normal hearing sensitivity, but none were rejected for obvious anatomical abnormalities.

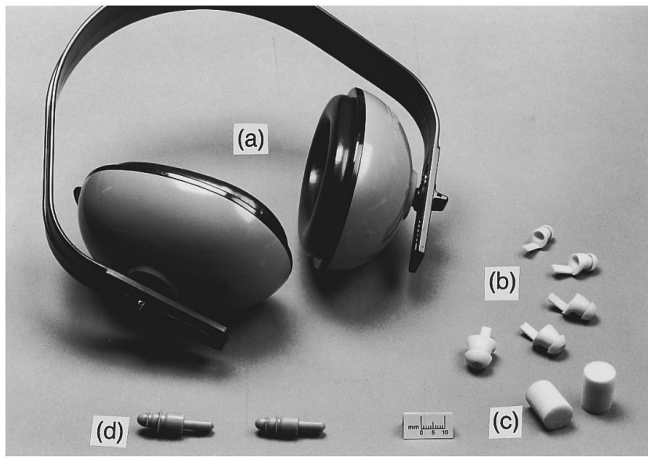


FIG. 1. Hearing protection devices tested in the S12/WG11 interlaboratory study: (a) UF-1 earmuff, (b) V-51R earplugs, (c) E-A-R Plugs, (d) EP100 earplugs.

C. Hearing protectors

Figure 1 is a photograph of the three types of earplugs and one earmuff which were used in the study. The E-A-R Plug[®], from Cabot Safety Corporation, is a slow-recovery foam plug available in a single size; the plug is compressed by rolling it into a thin, crease-free cylinder before insertion into the ear canal. These plugs may be considered to be disposable; therefore, a new pair of foam plugs was used for each occluded threshold test. The EP100 plug, from Willson Safety Products, is a premolded, double-flange plug that was available to subjects in two sizes. Subjects were given a pack containing both sizes from which they selected plugs at each fitting. The same pack of plugs was reused by a subject for the duration of the experiment. The V-51R plug, from PlasMed, Inc., is a premolded single-flange plug that was available to subjects in five sizes: extra large, large, medium, small, and extra small. The V-51R plugs used in this study were not color coded by size. As with the EP100, each subject was provided with a complete set of all sizes of the plugs, and the same pack of plugs was reused by a subject for the duration of the study. The UF-1 earmuff was available in one size and was worn in the standard band-over-the-head configuration.

Each laboratory was issued two numbered, identical earmuffs to use for its 24 subjects. Each subject always wore the same numbered device for all fittings. The cushions of the muffs were cleaned between subjects and were inspected daily by the experimenter for damage or excessive wear. Headband force measurements were performed on the earmuffs before and after the study to ensure that substantial alterations to the earmuffs did not occur in the course of the testing. All earmuff samples were initially obtained and measured for band force as received, by E•A•RCAL laboratory. The two samples sent to each laboratory were re-evaluated for band force both as received (pre-test), and subsequent to the complete set of REAT evaluations (post-test). All samples were returned to E•A•RCAL laboratory at the conclusion of the experiment, when they were measured one more time.

D. Apparatus

Testing was carried out at each of the four laboratories in test facilities meeting the requirements of ANSI S12.6-1984, with some modifications as noted below.

(1) The required test frequencies were 125, 250, 500, 1000, 2000, 4000, and 8000 Hz [omitting 3150 and 6300 Hz based on work of Berger and Rowland (1989)].

(2) A method for observing the subject during testing was required: either a viewing window into the test chamber or a video camera with remote monitor.

(3) A fitting noise was used in the IUF condition. The noise was a broadband random noise without audible tonal components, presented at an overall A-weighted sound pressure level of 65 dB. The spectral variation was required to be within a range of 20 dB from 100–8000 Hz. The sound pressure level in the 1/3 octave bands from 100–250 Hz was required to exceed the level in the 1000-Hz band by at least 15 dB.

E. Fitting instruction conditions

Two distinctly different fitting instruction conditions were compared: Subject Fit (SF) and Informed User Fit (IUF). Certain verbal directions to subjects for each fitting instruction condition were specified in the protocol and were read verbatim by the experimenters. To prevent interference with HPD fitting, subjects were not allowed to wear ear jewelry or eyeglasses.

1. SF Condition

Prior to entering the test chamber, subjects were given the hearing protector in the packaging in which it was sold along with the manufacturer's written fitting instructions. When applicable, both the individual package directions and the dispensing box directions were provided to subjects. In the SF procedure, the subject was initially instructed by the experimenter as follows:

"The purpose of this test is to estimate the noise reduction that you would be likely to obtain while wearing this hearing protector in a noisy environment. Please read the instructions and fit and adjust the hearing protector to the best of your ability. I am not allowed to assist you in that process."

For the hearing protectors that were available in multiple sizes (EP100 and V-51R), one pair of each size was placed before the subject and the following instructions were given:

"Please try these protectors on to find the size that is best for you. This may be different for each of your ears. Begin by trying a middle or regular size and then based on the looseness or tightness of the fit, and any guidance provided in the manufacturer's instructions, proceed to larger or smaller sizes as needed."

Subjects were given a maximum of five minutes to complete the fitting practice. Once satisfied with the fit, the subject removed the protector and entered the test chamber. In the chamber, subjects waited during a 2-min accommodation period before the first threshold measurement, either unoccluded or occluded. Just prior to each occluded threshold

measurement the following instructions were given to the subjects:

“After I leave the chamber, please put on the hearing protector in the way you have just learned. Refer to the manufacturer’s instructions as needed. Once you indicate that you have completed fitting the protector, the test will begin, and you may not touch or adjust the protector until you are asked to remove it at the end of the test. In the case of earplugs, if the device falls out of your ear during the test, please signal me. Throughout the test I will be able to observe you through the window [or using the TV camera].”

The experimenter was not present in the chamber while the subject fitted the HPDs prior to testing, and no assistance or additional explanations were permitted. Once the subject indicated that fitting was complete, testing began. For HPDs that require time to expand or conform to fit the ear canal or circumaural region (such as the E-A-R plug used in this experiment), the protocol for both SF and IUF conditions specified a 2-min waiting period after the subject fitted the device before testing began. After a test had begun, regardless of the adequacy of the fit of the protector, the data were accepted. If an earplug fell out of the ear during the test, the experimenter was to stop the test, enter the room, hand the subject the instructions and the earplugs and ask him/her to reinsert the devices for a retest. The protocol specified that if an earplug fell out a second time, testing would be stopped, anatomical measurements obtained, and the subject dismissed from the study. [However, fall-outs did not occur during the study.] Subjects remained in the chamber for a complete session consisting of two open and two occluded thresholds with the same HPD. Subjects were reminded that it was permissible to review the written instructions prior to each refit of the protector, but verbal or physical assistance from the experimenter was prohibited.

2. IUF condition

In the IUF procedure, the subject initially was instructed by the experimenter as follows:

“The purpose of this test is to estimate the noise reduction that you would be likely to obtain while wearing this hearing protector in a noisy environment.”

The experimenter continued by demonstrating on himself or herself the fitting of the HPD according to the manufacturer’s written instructions, then saying:

“Please carefully note the position of the earplugs in my ears” or

“Please carefully note the position of the earmuffs around my ears and on my head.”

The experimenter instructed each subject on how to size and fit the various protectors by using the manufacturer’s written instructions and by offering verbal feedback to the subject as the fitting progressed. Experimenters were not per-

mitted to augment inadequate manufacturer’s instructions or change inaccurate instructions. They were permitted to visually inspect a subject’s fit of a protector and offer an opinion as to whether the fit was adequate, and they could suggest that the subject remove a protector and try again, but they were not permitted to physically touch the protector or the subject. A fitting noise was presented to the subject to assist in selecting the proper size device and in obtaining an optimal fit of the protector. The fitting and practice session could last up to 8 min.

Following practice, the subject entered the test chamber for a 2-minute accommodation period prior to threshold measurements. The fitting noise was presented prior to each occluded threshold to aid the subject in achieving a good fit of the protector. Prior to the first occluded threshold measurement, the experimenter entered the chamber to give the following instructions:

“After I leave the chamber and have turned on the fitting noise, please put on the protector in the way you have just learned. Refer to the manufacturer’s instructions as needed. Adjust the hearing protector, while listening to the noise, so that it blocks the most sound from your ears but is still reasonably comfortable. By ‘blocking the most sound’ I mean that the noise you hear under the protector should be reduced to a minimum. By ‘reasonably comfortable’ I mean that you should consider your ability to wear the hearing protector for normal use while engaged in noisy occupational or recreational activities. Once you indicate that you have completed fitting of the protector, the noise will be turned off, the test will begin, and you may not touch or adjust the hearing protector until you are asked to remove it at the end of the test. In the case of earplugs, if the device falls out of your ear during the test, please signal me. Throughout the test I will be able to observe you through the window [or using the TV camera].”

The experimenter could not offer additional assistance after the subject entered the chamber, but the subject was free to consult the manufacturers’ instructions.

F. Optional test of muffs with glasses

Three of the four laboratories included one additional visit in which attenuation was assessed in two trials each for the SF and the IUF protocols for the combination of the UF-1 earmuff worn together with safety glasses (Aerosite model from Cabot Safety). Because workers may routinely wear corrective glasses or safety glasses with earmuffs, these additional data were collected to assess the effect of eyewear on attenuation. The Aerosite glasses were provided in two frame sizes, and the subject chose the preferred size.

G. Anthropometric measurements

At the conclusion of the last test session, anthropometric data were collected for all subjects in a standardized manner. The measurements included ear canal size, pinna length and breadth, tragus width of the head, and head height from the

earcanal to the top of the head. When these measurements were completed, subjects were thanked for their participation and dismissed.

III. RESULTS

A. Tabular results

Table II includes the means and standard deviations for the attenuation measurements on each HPD by lab and fitting instruction condition. The attenuation measurements from the two separate trials per HPD per visit were averaged before the standard deviations were calculated; therefore, the within-lab standard deviations shown are based on 24 measurements per condition. Table III includes similar data for the optional final test of earmuffs plus safety glasses.

B. Data analysis

The overall experiment had two between-subjects factors: laboratory (E•A•RCAL, NIOSH, USAARL, and WPAFB), and the order in which each HPD was tested by each subject (first, second, third, or fourth). There were three within-subjects factors: fitting instruction (subject fit and informed user fit), practice (first versus second repetition of testing with fitting instruction, i.e., SF1 versus SF2, or IUF1 versus IUF2), and HPD (E-A-R, EP100, UF-1, and V-51R).

Because differences in attenuation between HPDs were expected and were of no real interest for the purpose of this study, the data for each HPD were analyzed separately. Laboratory and subject were treated as random variables because the levels of these variables were viewed as samples from larger populations, and it was desired to make inferences beyond the particular laboratories and subjects chosen for this experiment. Fitting instruction and practice were treated as fixed variables because the levels of these variables were considered independent. Quasi F ratios (Winer, 1971, p. 375) were used to test some of the random effects because examination of the expected mean squares revealed that appropriate F ratios could not be calculated with individual mean squares. Summary results of repeated-measures analyses of variance performed at each test frequency for each HPD are shown as Tables IV–VII, with significance levels reported at $p < 0.05$.

The analyses summarized in Tables IV–VII show a significant main effect of subject (within lab \times order) for all devices, as well as a significant interaction of subject by fit (within lab \times order) for all the earplug devices. The dominant effect of subject variability has implications for the numbers of subjects required for HPD test protocols, as covered in Sec. V.

The effects of fitting instruction and practice can be seen in Fig. 2. Shown in Fig. 2, in a separate panel for each HPD, are the means of attenuation by frequency, averaged across labs, for each level of practice within fitting instruction (SF1, SF2, IUF1, and IUF2). For the earmuff, fitting instruction did not affect the mean attenuation achieved to any practically important degree, although the difference reached statistical significance at 1000 Hz and higher frequencies. For the earplugs, in contrast, fitting instruction had a large impact on attenuation: Most IUF means were at least 5 dB higher than

the corresponding SF means. The effect of fitting instruction reached statistical significance at all test frequencies for the V-51R and EP100 earplugs. For the E-A-R Plug, which yielded larger standard deviations of attenuation, the effect of fitting instruction reached significance only at 2000 and 8000 Hz (where the differences by fitting instruction were actually smaller). The effect of practice did not reach significance at any test frequency for any device.

The effect of laboratory can be seen in Figs. 3–6. In a separate figure for each device, the means and standard deviations of attenuation are shown in a separate panel for each fitting instruction condition (collapsed across practice, since it was not a significant effect). The effect of lab did not attain significance at any test frequency for the V-51R earplug or at any frequency except 8000 Hz for the EP100 earplug. For the E-A-R Plug and the UF-1 earmuff, the lab effect was significant at 1000 Hz and above. For the E-A-R Plug, there was a significant interaction of lab and fitting instruction at 125 and 500 Hz.

The effect of order was not significant except at 250 Hz for the UF-1 earmuff. However, because any potential order effects might limit the number of times subjects could be used in the HPD test protocol which S12/WG11 will develop, additional analyses were performed to look for order effects in SF1 data alone. In SF1, if subjects who tested a given device fourth were found to achieve higher attenuation than those who tested it first, this would indicate a need to limit the number of times a subject could be used. ANOVAs were performed separately at each test frequency for each device, using the effects of lab, order, subject (within lab \times order), and the lab \times order interaction. The effect of order attained statistical significance only at a single frequency for a single device (125 Hz for the UF-1 earmuff). Therefore, using subjects at least four times with four very different types of HPDs appears to be acceptable.

C. Interlaboratory variability by fitting instruction condition

The spread of attenuation results across labs by fitting instruction was of great interest to S12/WG11: Would interlab results be more similar in SF or in IUF? The single-lab means of attenuation by HPD for the SF and IUF fitting instruction conditions (collapsed across practice) were examined to determine interlab ranges. Shown in Fig. 7 are the ranges across labs in the means of attenuation for the SF and IUF conditions. For earplugs, the spread of results across labs is smaller in SF than in IUF, indicating less interlab variability in the SF fitting instruction condition. For the earmuff there is little difference.

In a similar fashion, the single-lab standard deviations of attenuation by HPD for the SF and IUF fitting instruction conditions (collapsed across practice) were examined and interlab ranges were determined. The ranges of standard deviations on attenuation across labs are plotted as Fig. 8. For all devices except the EP100 earplug, the ranges for SF and IUF interweave, indicating no systematic difference in variability by fitting instruction. For the EP100, there tended to be less interlab variability in standard deviations of attenuation in the SF condition than in the IUF condition.

TABLE II. Mean and standard deviation (s.d.) of attenuation by test frequency for each HPD, by fitting instruction condition and lab. The s.d. values are based on 24 observations; results of two trials per subject were averaged prior to summing across subjects.

CONDITION	TEST LAB	DATA STATISTIC	V-5IR PLUG						EPI00 PLUG						E-A-R PLUG						UF-1 MUFF									
			125		250		500		1000		2000		4000		8000		125		250		500		1000		2000		4000		8000	
			Attenuation Data, dB, by Frequency, Hz						Attenuation Data, dB, by Frequency, Hz						Attenuation Data, dB, by Frequency, Hz						Attenuation Data, dB, by Frequency, Hz									
SUBJECT FIT 1	EARCAL	Mean	14.4	13.7	14.4	16.4	24.7	25.9	19.1	18.0	17.0	18.8	20.9	29.1	34.2	32.3	22.5	23.0	25.7	26.8	33.6	41.6	41.2	7.6	14.9	21.6	31.3	32.6	36.9	36.3
		s.d.	10.5	10.8	10.1	10.0	9.4	8.7	12.7	9.9	10.4	11.0	9.0	7.9	8.7	10.8	4.8	5.3	6.4	5.8	2.5	3.2	5.8	3.5	3.1	2.7	3.6	3.5	3.7	4.1
USAARL	NIOSH	Mean	11.0	10.5	10.5	11.9	17.6	19.0	15.6	13.4	13.3	13.8	15.4	20.5	28.1	23.9	18.8	18.9	20.3	19.5	27.1	34.9	33.4	7.4	14.0	19.0	25.6	28.6	33.3	32.8
		s.d.	10.1	9.5	9.9	8.8	9.4	6.6	9.7	10.0	9.8	10.6	9.5	8.6	10.4	14.4	7.8	6.5	6.3	5.3	4.5	7.0	9.1	3.4	3.3	2.9	3.9	3.3	4.4	4.9
WPAFB	NIOSH	Mean	11.2	10.7	10.9	13.6	21.8	22.6	16.8	13.2	12.5	12.4	14.7	22.7	28.4	23.2	23.7	25.1	26.2	27.2	31.9	40.7	42.0	7.7	14.5	21.3	30.3	31.8	38.3	37.0
		s.d.	9.5	8.2	8.5	9.0	8.6	7.0	11.3	11.2	11.2	11.9	11.2	11.7	11.4	14.6	4.1	4.0	5.1	4.6	3.5	3.3	5.0	3.6	2.6	2.3	2.5	3.8	3.6	4.7
WPAFB	NIOSH	Mean	10.6	9.9	11.3	14.7	22.3	24.7	20.8	13.0	14.3	15.0	16.4	23.6	29.0	27.9	17.9	19.0	21.0	24.7	29.9	35.6	34.6	7.0	13.3	21.5	30.4	33.6	34.9	34.9
		s.d.	9.3	8.8	9.2	9.6	10.4	7.7	7.6	11.7	11.2	12.9	11.8	12.0	11.5	12.9	7.3	6.3	7.3	6.4	5.3	5.0	5.4	4.1	2.5	3.2	3.7	4.2	2.8	3.1
SUBJECT FIT 2	EARCAL	Mean	13.7	13.3	14.2	16.0	22.5	23.9	21.7	19.2	18.1	19.3	20.4	28.6	34.7	31.7	25.5	26.0	28.9	29.1	34.8	42.1	42.6	7.8	14.2	21.2	30.7	32.3	34.8	34.5
		s.d.	11.2	10.4	10.9	11.0	9.7	8.5	14.7	12.1	12.1	12.2	10.9	8.5	11.0	14.3	7.5	7.5	8.1	7.3	3.3	3.5	6.1	2.9	3.2	2.6	3.0	3.4	3.2	4.2
USAARL	NIOSH	Mean	10.5	9.7	10.9	12.8	19.2	20.0	17.2	13.3	13.5	14.2	16.2	22.0	27.9	23.7	19.5	19.3	21.1	20.5	27.4	35.4	33.7	7.6	14.4	20.1	26.4	29.6	34.4	33.4
		s.d.	9.3	9.3	9.2	9.6	9.1	6.8	10.5	11.3	11.6	12.1	10.4	9.6	11.0	14.1	8.3	6.6	6.6	5.9	5.5	7.3	7.9	2.7	2.5	2.6	3.9	3.5	3.5	4.7
WPAFB	NIOSH	Mean	11.4	10.7	10.5	14.4	22.5	23.3	18.7	14.6	14.5	15.2	17.6	24.9	29.8	25.3	21.5	22.4	24.9	26.1	30.8	39.9	40.5	7.1	13.2	20.3	29.3	31.4	38.0	35.3
		s.d.	10.2	9.6	9.1	8.7	8.5	7.1	10.8	12.9	12.7	14.0	11.8	10.8	11.4	14.5	6.9	5.7	6.8	5.5	4.1	6.2	5.3	3.4	3.1	3.3	3.1	3.9	4.2	6.0
WPAFB	NIOSH	Mean	8.9	8.7	9.7	13.0	19.9	23.4	19.3	12.8	13.1	14.6	18.3	23.7	28.9	27.8	22.0	22.5	25.3	27.7	32.3	37.3	38.0	7.1	13.5	20.6	30.0	33.5	34.3	34.5
		s.d.	9.5	8.9	9.6	10.7	10.2	8.0	6.1	11.1	10.9	12.5	11.9	11.7	11.7	12.4	6.9	7.8	8.7	7.9	4.3	5.4	4.0	2.9	2.7	3.0	3.3	4.6	3.0	3.3
INFORMED USER FIT 1	EARCAL	Mean	21.8	21.1	22.2	22.3	29.6	30.5	33.8	26.2	24.6	26.4	26.6	34.5	40.9	40.9	34.4	35.9	38.7	37.7	36.8	43.7	46.6	9.8	14.8	21.9	32.1	33.0	36.6	36.9
		s.d.	9.2	9.1	8.6	7.4	5.9	7.0	10.1	6.1	6.9	6.8	5.2	4.3	6.5	7.3	6.1	5.9	5.9	5.0	2.9	2.9	3.8	3.5	2.8	2.2	2.4	2.6	3.7	3.7
USAARL	NIOSH	Mean	14.6	14.2	15.1	16.1	22.6	24.0	23.5	17.5	16.4	17.3	18.3	23.7	31.0	27.8	21.1	20.7	22.4	22.6	28.7	37.1	36.2	7.5	13.7	19.2	26.9	29.6	35.0	33.8
		s.d.	9.1	9.5	9.7	8.9	8.8	7.4	12.5	11.6	11.0	11.2	10.3	9.3	11.8	14.7	6.6	6.5	6.3	4.7	4.6	5.3	7.4	3.1	3.0	2.6	3.7	3.5	3.5	3.2
WPAFB	NIOSH	Mean	16.9	16.0	16.9	19.8	26.0	26.6	25.0	17.5	16.5	17.1	19.8	27.0	33.0	29.7	26.0	26.6	28.8	29.5	32.7	40.8	42.6	8.7	13.2	20.4	30.1	32.8	39.1	37.5
		s.d.	8.7	8.2	7.6	7.5	6.8	5.1	10.5	9.9	9.7	9.3	8.6	8.6	9.2	10.5	5.6	6.1	6.6	6.1	5.0	5.8	4.9	3.5	3.2	2.7	3.1	2.5	3.1	4.8
WPAFB	NIOSH	Mean	15.2	15.1	16.0	20.5	26.2	28.0	24.7	20.3	20.3	22.8	25.0	31.1	37.2	36.7	26.9	26.8	30.3	31.3	33.4	39.1	39.8	7.4	13.5	20.7	30.9	34.2	35.5	35.5
		s.d.	8.8	8.0	7.5	7.4	7.4	6.2	6.2	8.0	8.2	8.6	6.8	8.0	6.5	7.7	7.0	7.7	8.0	6.2	3.6	3.3	3.3	3.7	2.2	2.1	4.0	2.2	2.3	2.1
INFORMED USER FIT 2	EARCAL	Mean	22.7	22.9	22.5	23.5	30.0	30.8	35.4	25.3	24.7	25.7	25.8	33.2	40.1	39.2	33.7	36.2	39.2	38.2	36.8	44.1	46.7	9.7	14.9	22.3	31.3	32.5	36.2	36.4
		s.d.	7.2	7.1	7.0	6.8	5.4	5.9	9.6	8.0	8.1	8.7	7.1	6.7	7.4	10.9	5.7	6.2	5.9	4.4	2.5	3.0	3.6	4.0	3.0	3.0	3.5	3.8	3.4	4.0
USAARL	NIOSH	Mean	13.2	12.9	13.9	17.4	22.2	22.7	22.1	15.8	16.0	16.7	18.9	25.0	31.8	27.7	21.7	20.5	22.4	21.7	29.0	36.4	35.5	7.7	14.3	20.1	27.3	29.5	34.8	34.0
		s.d.	8.4	8.5	8.8	8.4	8.4	6.2	12.8	11.2	11.0	11.3	9.8	9.1	10.5	13.2	7.8	7.3	7.2	6.5	4.8	5.0	6.6	2.8	2.7	2.6	4.0	3.2	4.4	4.6
WPAFB	NIOSH	Mean	16.3	15.9	15.0	19.1	25.8	26.8	24.6	19.8	19.4	19.2	20.8	26.7	35.2	30.5	24.2	25.5	28.3	28.8	32.4	40.2	41.5	8.8	13.7	21.5	30.3	32.0	39.0	37.3
		s.d.	9.1	8.8	8.9	8.1	8.7	6.8	10.5	11.9	11.1	11.4	9.9	8.9	9.6	11.6	8.0	7.5	7.0	7.2	4.6	4.8	4.8	4.0	2.8	2.7	3.0	3.3	3.1	4.7
WPAFB	NIOSH	Mean	13.5	13.2	14.9	20.3	26.4	27.8	24.0	21.2	21.3	24.0	25.1	30.5	37.4	37.1	26.0	29.4	32.2	32.8	34.0	39.3	39.6	6.9	13.5	20.4	30.5	34.0	35.9	35.4
		s.d.	7.9	7.3	7.5	6.5	7.1	6.2	6.5	10.8	9.1	9.9	8.6	7.3	6.5	7.7	7.3	7.5	7.9	6.1	3.2	3.0	3.4	2.8	3.0	3.1	3.8	2.7	3.1	3.0

TABLE III. Mean and standard deviation (s.d.) of attenuation by test frequency for the UF-1 earmuff worn with Aerosite safety glasses, by fitting instruction condition and lab. The s.d. values are based on 24 observations (two trials averaged) except in the EARCAL lab, where only 23 subjects participated.

TEST CONDITION	TEST LAB	DATA STATISTIC	UF-1 MUFF PLUS SAFETY GLASSES						
			Attenuation Data, dB, by Frequency, Hz						
			125	250	500	1000	2000	4000	8000
SUBJECT FIT	EARCAL	Mean	4.7	7.8	16.4	25.3	28.6	30.3	25.9
		s.d.	3.4	4.0	3.5	3.7	4.0	4.3	5.9
	USAARL	Mean	2.8	5.5	13.4	21.6	25.5	28.7	22.5
		s.d.	3.3	4.2	3.6	2.6	4.2	4.3	5.9
	NIOSH	Mean	4.5	6.0	14.9	24.3	28.7	32.3	28.7
		s.d.	4.2	4.6	3.7	4.1	4.9	5.7	7.2
INFORMED USER FIT	EARCAL	Mean	6.8	9.6	18.0	27.4	30.5	31.8	28.1
		s.d.	4.3	3.7	2.9	3.7	4.5	4.5	5.0
	USAARL	Mean	4.2	6.1	13.5	22.1	26.2	28.5	24.2
		s.d.	3.3	3.3	2.8	3.2	4.6	4.4	5.5
	NIOSH	Mean	4.7	6.6	14.9	25.1	28.5	32.6	29.7
		s.d.	4.3	4.7	3.4	4.2	4.9	5.9	7.5

D. Effect of safety glasses on earmuff attenuation and interlab variability

Wearing safety glasses in combination with earmuffs reduced the mean attenuation achieved by about 5 dB or more across all test frequencies, as shown in Fig. 9 for data averaged across the three participating labs. However, for the development of the protocol to estimate real-world protection, the absolute effect of glasses on attenuation was of less interest to S12/WG11 than whether glasses would increase interlaboratory variability to unacceptable levels. Presented in Fig. 10 are the means and standard deviations of attenua-

tion by lab for the tests of muffs plus glasses. (Note that at E•A•RCAL lab, only 23 of 24 subjects participated in the glasses test.) When Figs. 6 and 10 are compared, it is evident that the within-laboratory standard deviations do increase with glasses, but only slightly. The standard deviations for muffs plus glasses are still lower than for any earplug tested.

E. Earmuff band force measurements

The pre-test and post-test measurements of band force for the earmuffs tested in each laboratory are presented in

TABLE IV. Summary of quasi F ratios from ANOVAs conducted separately at each test frequency for the V-51R plug. Asterisks indicate values reaching statistical significance at $p < 0.05$ or less.

SOURCE OF VARIANCE	F RATIOS BY TEST FREQUENCY, Hz						
	125	250	500	1000	2000	4000	8000
BETWEEN SUBJECTS							
Lab	2.13	2.19	1.89	1.10	1.71	3.91	1.67
Order	0.66	0.52	0.49	0.25	0.09	0.36	0.22
Lab×Order	1.40	1.44	1.76	1.96	2.05	1.54	0.84
Subject (Lab×Order)	3.44*	3.58*	3.99*	4.42*	4.39*	3.33*	4.44*
WITHIN SUBJECTS							
Fitting Instruction	24.23*	26.75*	37.31*	114.31*	72.12*	76.22*	13.71*
Lab×Fit Instr	2.39	4.76	2.60	1.00	0.44	0.26	5.20
Order×Fit Instr	3.24	3.98*	4.73*	3.89*	2.66	0.69	1.12
Lab×Order×Fit Instr	0.99	0.63	0.85	0.68	1.09	2.20	1.31
Subject×Fit (Lab×Order)	2.44*	2.30*	1.94*	2.26*	3.26*	2.61*	2.71*
Practice	3.29	0.91	4.55	0.17	0.47	1.05	0.46
Lab×Practice	1.70	1.68	1.66	0.99	0.70	0.80	2.62
Order×Practice	1.31	0.80	4.45*	0.83	1.95	1.72	2.27
Lab×Order×Practice	0.65	0.91	1.08	1.92	0.78	0.44	0.32
Subject×Practice (Lab×Order)	0.89	0.83	0.71	0.79	0.87	1.36	1.27
Fit Instr×Practice	0.00	0.21	0.75	0.48	0.23	0.01	2.34
Lab×Fit Instr×Practice	0.26	0.40	0.51	0.92	2.36	1.27	0.41
Order×Fit Instr×Practice	0.40	0.28	0.30	0.57	0.38	0.22	0.11
Lab×Order×Fit Instr×Practice	1.18	1.00	0.51	0.64	0.98	1.59	1.80

TABLE V. Summary of quasi F ratios from ANOVAs conducted separately at each test frequency for the EP100 plug. Asterisks indicate values reaching statistical significance at $p < 0.05$ or less.

SOURCE OF VARIANCE	F RATIOS BY TEST FREQUENCY, Hz						
	125	250	500	1000	2000	4000	8000
BETWEEN SUBJECTS							
Lab	3.01	2.52	2.39	2.25	3.07	3.65	5.63*
Order	1.77	2.27	1.55	1.11	0.46	2.19	1.13
Lab×Order	0.93	0.82	1.01	1.07	1.74	0.86	0.78
Subject (Lab×Order)	3.67*	3.87*	3.92*	4.24*	3.98*	3.93*	4.29*
WITHIN SUBJECTS							
Fitting Instruction	29.88*	26.52*	20.78*	23.29*	22.94*	29.54*	34.47*
Lab×Fit Instr	1.73	2.34	2.30	3.69	7.66	3.50	1.61
Order×Fit Instr	0.70	1.09	0.73	0.94	1.53	1.49	0.79
Lab×Order×Fit Instr	1.06	0.90	1.20	0.68	0.38	0.65	1.28
Subject×Fit (Lab×Order)	2.39*	2.26*	2.75*	2.47*	2.63*	2.73*	2.09*
Practice	0.44	1.44	1.18	2.13	0.33	1.39	0.02
Lab×Practice	1.21	1.82	2.15	1.60	2.44	1.00	2.31
Order×Practice	1.80	3.79	3.37	1.40	1.87	1.50	2.71
Lab×Order×Practice	0.81	0.61	0.51	1.42	0.57	1.48	0.86
Subject×Practice (Lab×Order)	1.38	1.29	1.48*	1.14	1.32	1.21	0.95
Fit Instr×Practice	0.25	0.36	0.24	5.33	4.08	0.14	1.09
Lab×Fit Instr×Practice	1.08	0.70	0.67	0.64	0.63	0.74	0.22
Order×Fit Instr×Practice	2.89	2.70	6.52*	8.58*	3.28	3.57	2.62
Lab×Order×Fit Instr×Practice	0.54	0.62	0.53	0.35	0.53	0.36	0.44

Table VIII. The columns labeled as “initial” and “final” show the measurements of all earmuffs at E•A•RCAL lab prior to the experiment and after its conclusion. The “initial” data indicated that the new samples exhibited very little variation in band force, approximately 2%. However, the range in measured force across laboratories on both the pre-test and post-test samples was 1.3 and 1.7 N, respectively, or about 15%–18% of the average measured values. In general

there was a slight drop in force between the beginning and end of the multi-subject test series. By comparison, in a similar interlaboratory study (Poulsen, 1984) wherein three pairs of UF-1 earmuffs (same model as used in this study) were circulated among four laboratories, the range in measured band force across laboratories was 31%.

Although band force measurements can be used as a descriptive statistic, or as a measure of the change in earmuff

TABLE VI. Summary of quasi F ratios from ANOVAs conducted separately at each test frequency for the E-A-R Plug. Asterisks indicate values reaching statistical significance at $p < 0.05$ or less.

SOURCE OF VARIANCE	F RATIOS BY TEST FREQUENCY, Hz						
	125	250	500	1000	2000	4000	8000
BETWEEN SUBJECTS							
Lab	2.19	2.36	3.01	5.31*	13.80*	12.56*	12.28*
Order	0.53	0.72	0.82	0.71	0.33	1.89	0.10
Lab×Order	1.59	1.61	1.56	1.22	1.01	0.76	0.77
Subject (Lab×Order)	3.88*	3.40*	3.60*	3.07*	5.43*	4.45*	5.51*
WITHIN SUBJECTS							
Fitting Instruction	8.43	5.92	7.09	7.79	31.74*	9.29	11.35*
Lab×Fit Instr	10.52*	28.24	28.20*	67.30	0.97	2.79	1.98
Order×Fit Instr	0.78	1.03	0.45	0.69	1.62	0.82	0.18
Lab×Order×Fit Instr	0.91	0.55	0.61	0.42	0.99	0.95	1.60
Subject×Fit (Lab×Order)	1.49*	1.84*	1.87*	2.58*	1.82*	1.74*	1.59*
Practice	0.38	0.51	1.63	1.00	0.90	0.17	0.13
Lab×Practice	2.19	3.57	2.51	2.56	1.65	1.66	1.41
Order×Practice	1.73	0.38	0.50	0.08	0.07	1.79	0.55
Lab×Order×Practice	1.89	0.71	1.67	1.00	1.91	1.27	1.32
Subject×Practice (Lab×Order)	0.94	1.10	0.79	1.13	0.97	0.90	1.13
Fit Instr×Practice	2.42	0.48	2.79	5.25	0.96	2.84	2.71
Lab×Fit Instr×Practice	2.16	1.00	0.84	0.41	2.37	0.71	1.32
Order×Fit Instr×Practice	1.37	0.54	0.27	0.34	2.30	0.12	0.88
Lab×Order×Fit Instr×Practice	0.59	1.22	0.85	1.45	0.68	0.68	1.31

TABLE VII. Summary of quasi F ratios from ANOVAs conducted separately at each test frequency for the UF-1 earmuff. Asterisks indicate values reaching statistical significance at $p < 0.05$ or less.

SOURCE OF VARIANCE	F RATIOS BY TEST FREQUENCY, Hz						
	125	250	500	1000	2000	4000	8000
BETWEEN SUBJECTS							
Lab	1.27	4.60	4.98	16.56*	11.34*	14.66*	8.39*
Order	3.36	4.09*	1.74	2.65	0.38	1.94	3.62
Lab×Order	0.50	4.86	0.40	0.71	0.97	0.61	0.43
Subject (Lab×Order)	8.18*	7.61*	4.70*	5.56*	5.33*	9.76*	6.26*
WITHIN SUBJECTS							
Fitting Instruction	3.27	0.08	0.28	25.30*	26.81*	48.81*	50.34*
Lab×Fit Instr	9.18	0.33	1.20	0.55	0.46	0.16	0.34
Order×Fit Instr	0.79	0.39	1.51	2.59	0.78	0.90	0.32
Lab×Order×Fit Instr	0.93	0.22	0.84	0.71	0.60	1.07	1.04
Subject×Fit (Lab×Order)	1.22	2.70	1.06	1.55	1.59	0.97	1.05
Practice	0.12	0.00	0.07	0.68	0.46	0.74	1.83
Lab×Practice	0.94	0.70	1.56	2.28	1.61	1.10	1.64
Order×Practice	3.89*	0.62	1.03	3.98	0.95	1.12	1.01
Lab×Order×Practice	1.28	2.25	1.16	0.44	2.59	3.89	1.35
Subject×Practice (Lab×Order)	0.75	0.82	0.80	1.10	0.66	0.69	0.88
Fit Instr×Practice	0.05	2.16	3.26	0.18	3.44	0.38	2.64
Lab×Fit Instr×Practice	0.59	1.72	1.53	0.59	0.47	3.85*	0.82
Order×Fit Instr×Practice	0.59	1.25	0.13	0.41	2.12	0.21	0.56
Lab×Order×Fit Instr×Practice	0.71	0.68	0.79	1.59	0.63	0.46	0.71

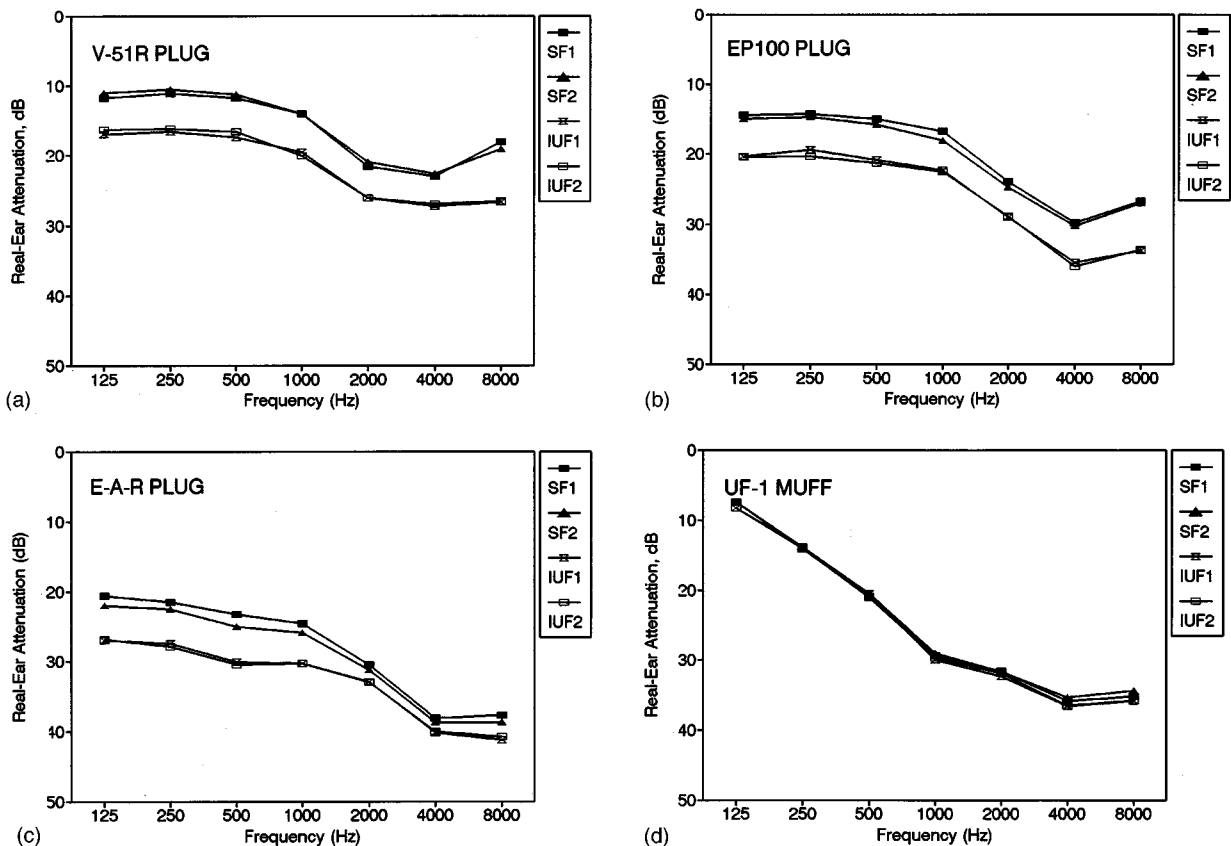


FIG. 2. Mean attenuations across four laboratories by fitting instruction and practice for the V-51R earplug (a), EP100 earplug (b), E-A-R Plug (c), and UF-1 earmuff (d).

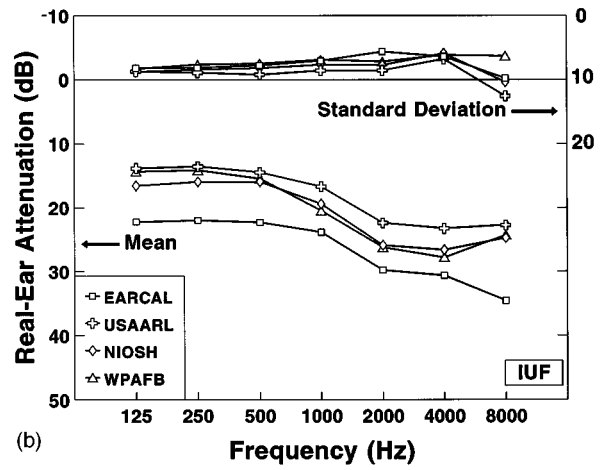
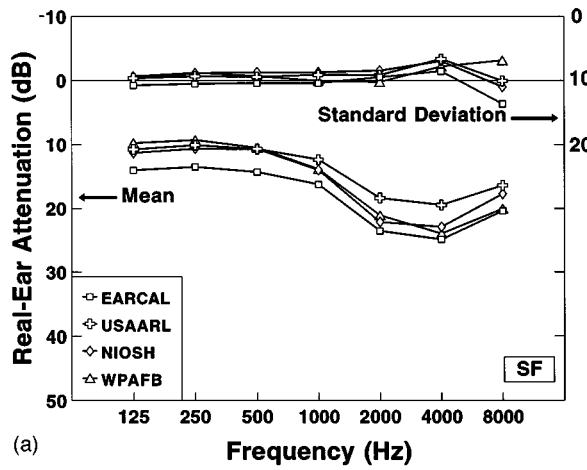


FIG. 3. Means and standard deviations of attenuation for the V-51R earplug by lab for SF (a) and IUF (b) fitting instruction conditions.

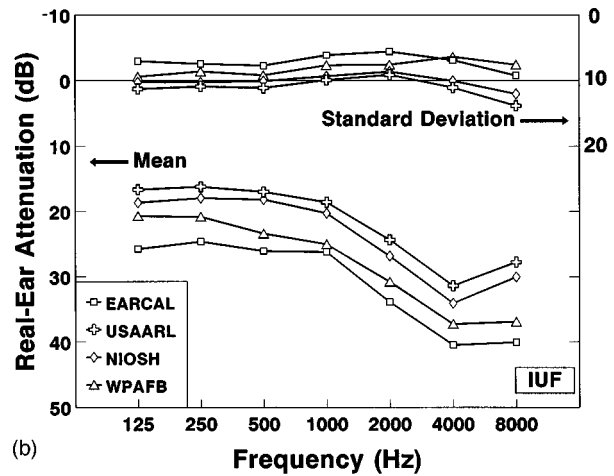
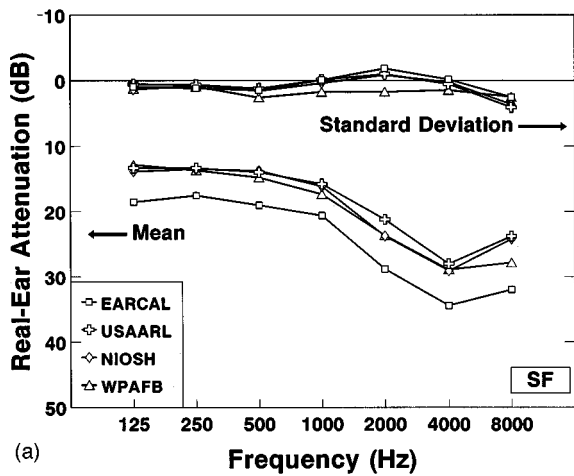


FIG. 4. Means and standard deviations of attenuation for the EP100 earplug by lab for SF (a) and IUF (b) fitting instruction conditions.

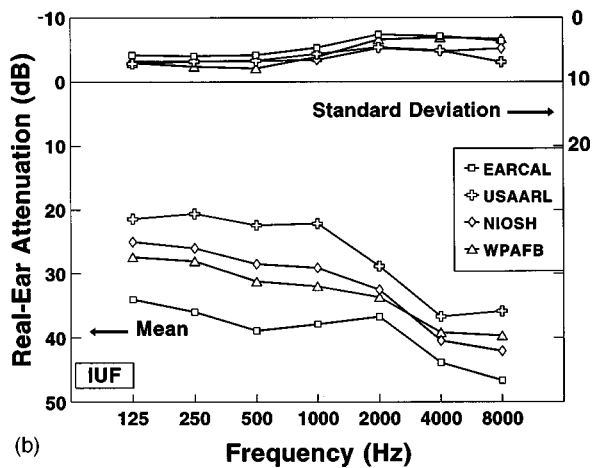
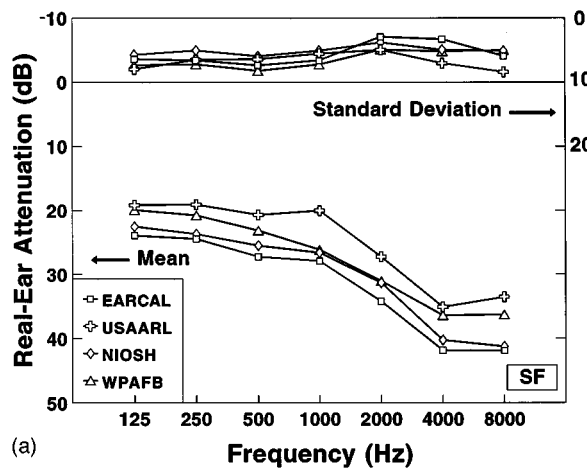


FIG. 5. Means and standard deviations of attenuation for the E-A-R Plug by lab for SF (a) and IUF (b) fitting instruction conditions.

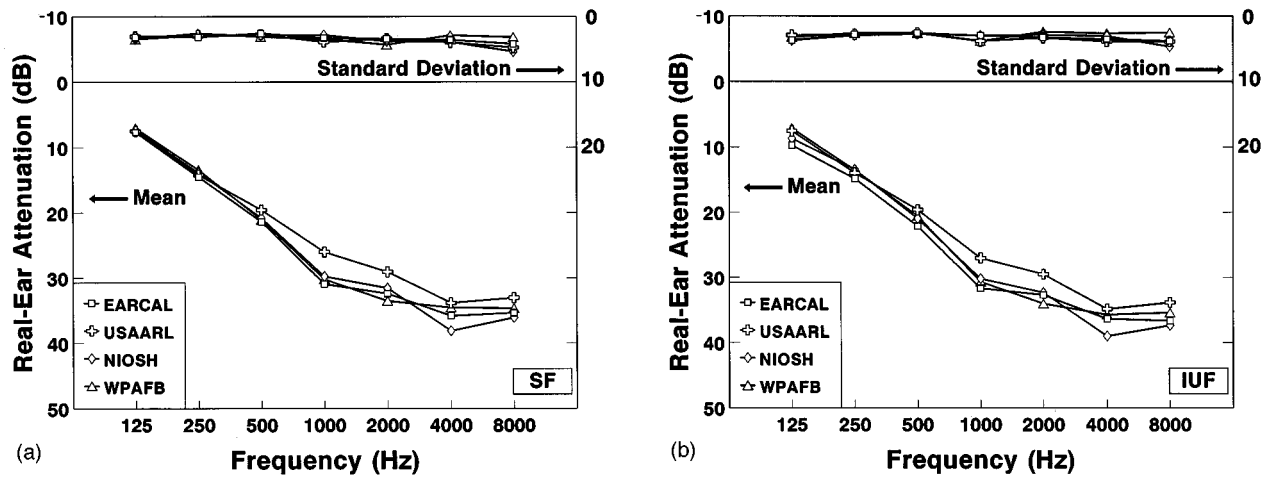


FIG. 6. Means and standard deviations of attenuation for the UF-1 earmuff by lab for SF (a) and IUF (b) fitting instruction conditions.

performance in a single laboratory, the apparent imprecision of force measurements across labs indicates that care should be taken in using them to set performance requirements at this time.

IV. COMPARISON OF S12/WG11 INTERLABORATORY VARIABILITY TO THAT OF OTHER STUDIES

The variability of attenuation values obtained in the S12/WG11 interlaboratory study was compared with that found in the following studies.

1. U.S. Air Force (AF) Survey (Johnson, 1976)

In 1976, Air Force personnel retrieved hearing protector attenuation and standard deviation data from different laboratories that had already completed contracted measurements on several hearing protectors. Data for two earplugs and one earmuff were collected from four laboratories. Attenuation measurements were accomplished by all laboratories in compliance with American National Standard Z24.22-1957 using discrete tone stimuli presented in anechoic spaces. Results of this comparison indicated that

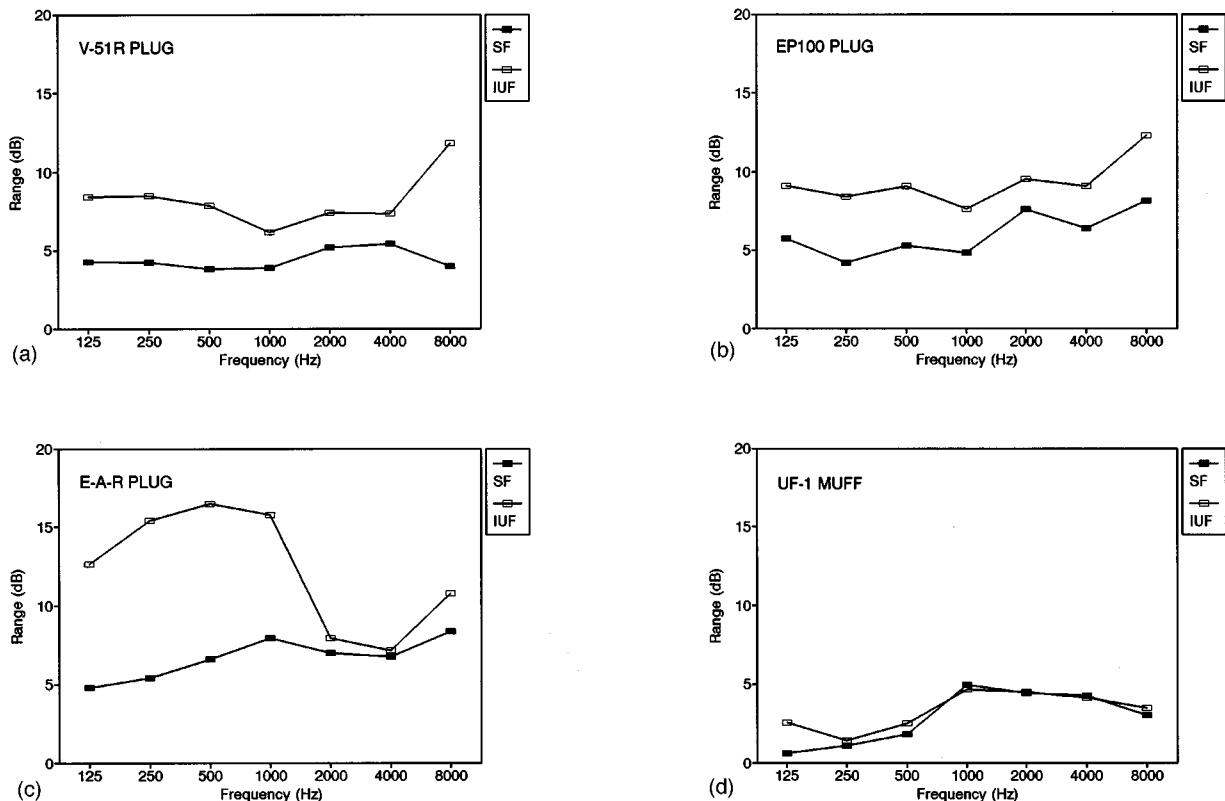


FIG. 7. Ranges of results from four laboratories for values of mean attenuation in the SF and IUF fitting instruction conditions (collapsed across practice) for the V-51R earplug (a), EP100 earplug (b), E-A-R Plug (c), and UF-1 earmuff (d).

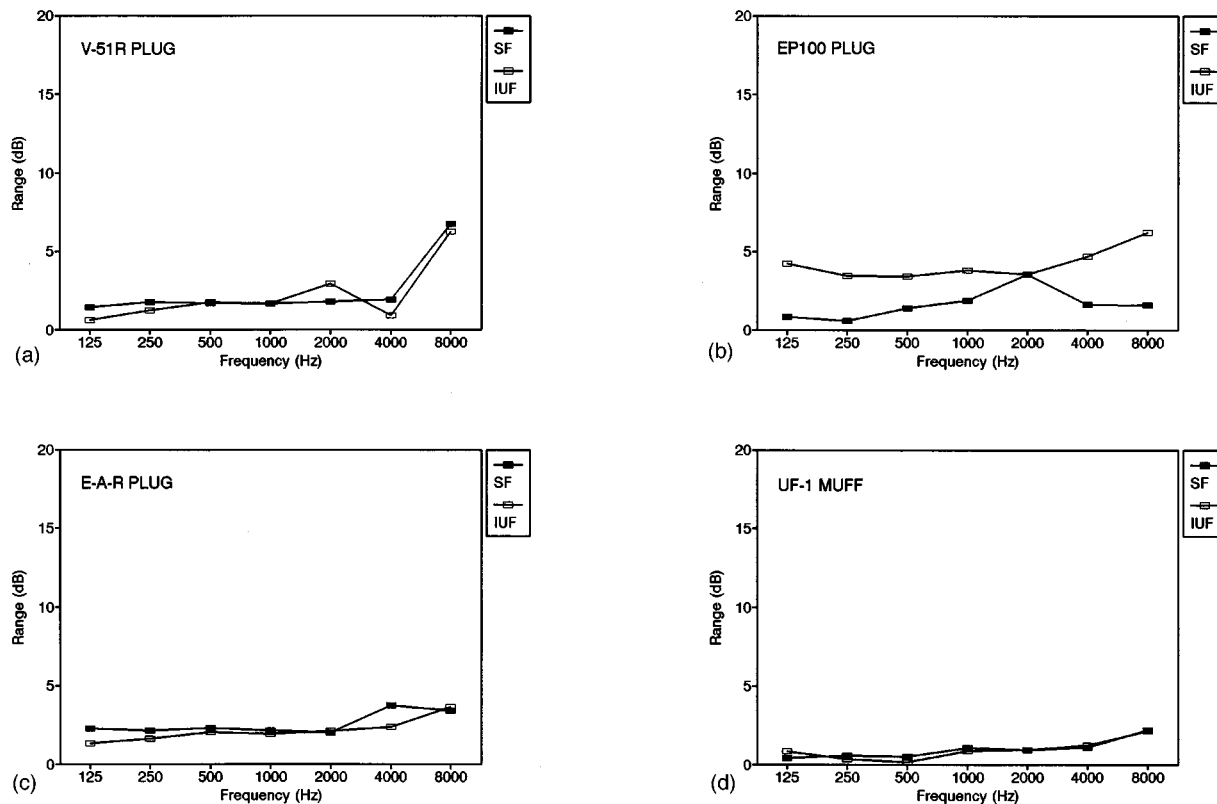


FIG. 8. Ranges of results from four laboratories for standard deviations of attenuation in the SF and IUF fitting instruction conditions (collapsed across practice) for the V-51R earplug (a), EP100 earplug (b), E-A-R Plug (c), and UF-1 earmuff (d).

interlaboratory differences ranged from 5–15 dB for earplugs and from 2–8 dB for the earmuff.

2. Environmental Protection Agency (EPA) Study (Berger *et al.*, 1982)

This first planned study of interlaboratory variability in the measurement of HPD attenuation was initiated by EPA to have four HPDs evaluated by seven labs in a round robin test program. Data from an eighth lab already on file with the EPA were also included in the analyses. The measurements were accomplished in accordance with EPA (1979) requirements and ANSI S3.19-1974. The results showed significant variability in both mean attenua-

tion and standard deviation values, suggesting that even the rank ordering of hearing protection devices on the basis of measured attenuation was not possible across laboratories. Primary sources of variance were associated with the proper fit of the hearing protection devices, subject selection and training, and treatment of the data.

3. European Economic Community (EEC) Study (Shipton, 1986)

Four earmuffs and the E-A-R foam earplug were evaluated by the National Physical Laboratory and five other labs in the United Kingdom under sponsorship of the EEC, in accordance with the procedures of ISO 4869:1981. Results are described together with those of the next study.

4. Nordic Round Robin Study (Poulsen, 1984)

Four earmuffs and two earplugs were evaluated by four laboratories following ISO 4869:1981. Considering both the EEC and Nordic studies together (since the same method was used), the results indicated that repeated measurements in the same laboratory showed a high level of reliability. For earplugs, the inter-subject variability was high, with the result that large differences between labs were not statistically significant. For earmuffs, inter-subject variability was lower, resulting in significant lab differences even though the data were in fairly good agreement.

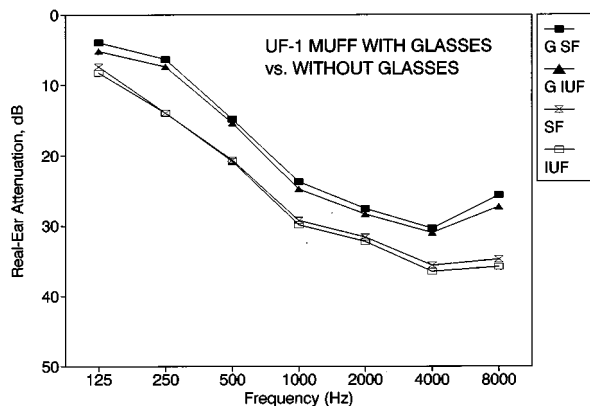


FIG. 9. Mean attenuation across three laboratories for the UF1 earmuff in fitting instruction conditions SF1 and IUF2 (no safety glasses worn) versus the final SF and IUF tests of the muff worn with safety glasses.

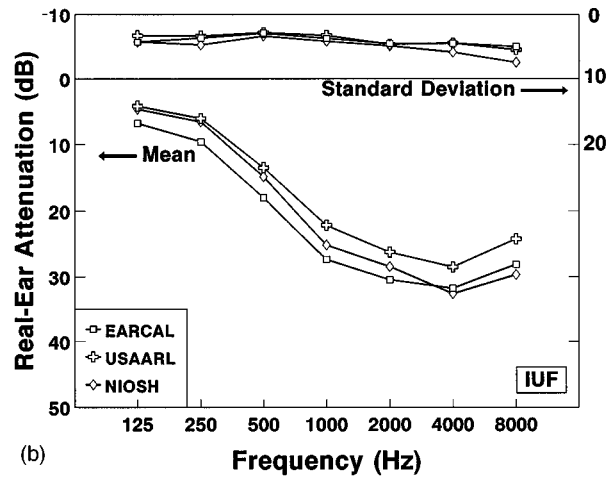
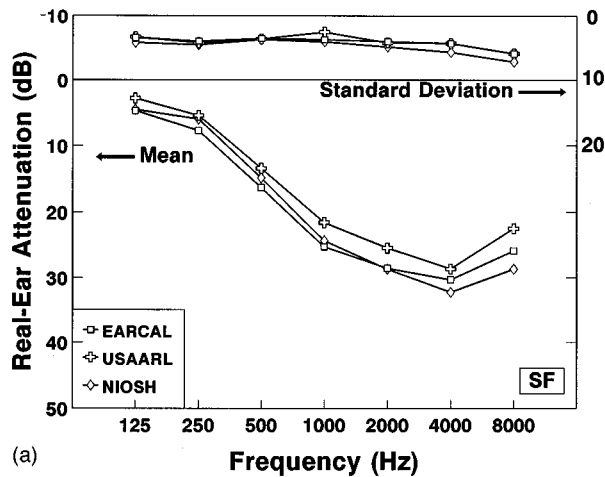


FIG. 10. Means and standard deviations of attenuation in each laboratory for the UF-1 earmuff worn with safety glasses in the SF (a) and IUF (b) fitting instruction conditions.

gency and variance data for these studies had been published. These data were examined for each of the earlier studies, plus the SF1 data from the current S12/WG11 study, to determine the correspondence among the values measured under the different conditions between 1976 and 1993. HPDs were selected for comparison on the basis of their inclusion among the various studies, with only the E-A-R foam earplug appearing in all studies. The only other devices appearing in more than one study were the V-51R earplug (three times), and the Optac Optigard and the Sonogard earmuffs (twice each). Consequently, the three data sets examined were for the E-A-R and V-51R earplugs and a group of four different earmuffs representing the different studies. These data were obtained using different protocols, different numbers of subjects (from 10 to 24), and different numbers of trials per subject per device (1, 2, or 3).

A. E-A-R Plug data

In order to illustrate the range of mean attenuation values obtained in each study, the lowest and highest mean attenuation values obtained by any participating lab were identified for each test frequency.

These low and high values for each study are plotted as Fig. 11, forming ranges with different degrees of overlap. The low values are similar for the S12/WG11 SF1 data and the EEC and Nordic studies, although their high values differ. The Nordic study exhibits the least interlab variability, as shown by the narrow range between low and high attenuation values. The AF survey and EPA data are comparable in terms of both low and high values, and these two studies showed greater attenuation, as evidenced by very little overlap between low values for EPA and AF survey data and high values for the other studies.

To focus better on the relative interlab variability for the E-A-R data from the five studies, the ranges of the means and the standard deviations of attenuation are presented as Fig. 12. The Nordic study achieved the smallest range of mean attenuation at six of seven test frequencies. The ranges of mean values for the S12/WG11 SF data are similar to the other studies at all frequencies, always showing values less than or equal to at least one of the other studies. The EEC study exhibited the largest interlaboratory ranges at most test

TABLE VIII. Band force in Newtons for UF-1 earmuffs as measured by each laboratory before and after the attenuation test series. All muff samples were measured at EARCAL lab before distribution to the other labs, and again after the end of the entire experiment.

TEST LAB	BAND FORCE, N, BY MEASUREMENT TIME			
	EARCAL initial	Pre-Test	Post-Test	EARCAL final
EARCAL	9.7	9.7	9.2	9.2
	9.8	9.8	9.3	9.3
USAARL	9.6	9.2	9.2	9.4
	9.8	9.4	9.4	9.5
NIOSH	9.7	na	8.2	9.6
	9.6	na	8.9	9.9
WPAFB	9.6	8.0	8.0	9.6
	9.7	8.9	8.9	9.8
Mean	9.7	9.2	8.9	9.5
s.d.	0.1	0.6	0.5	0.2

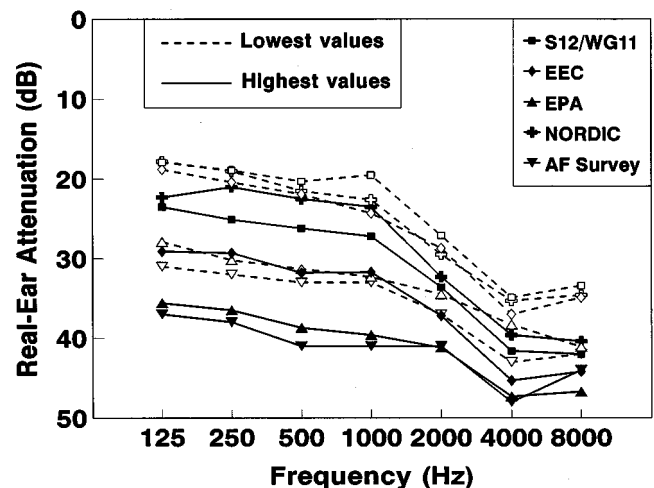
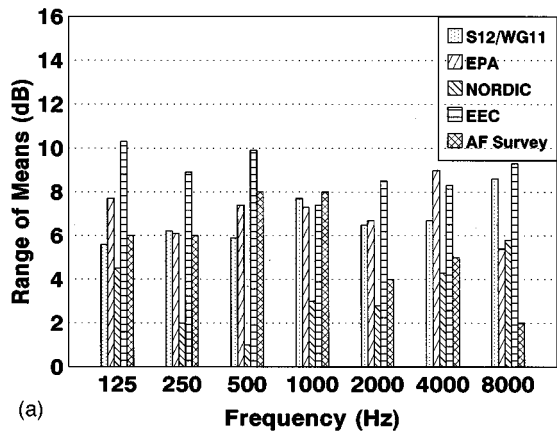
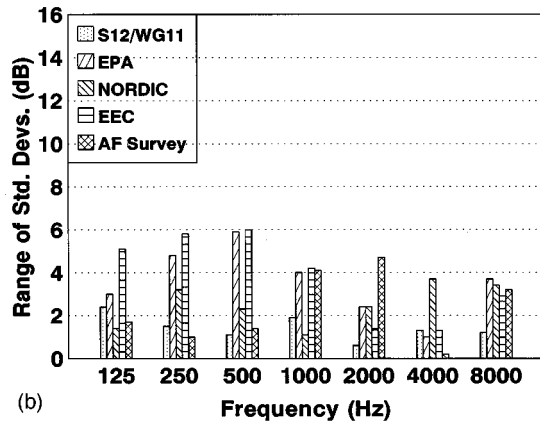


FIG. 11. Lowest and highest mean attenuation values obtained for the E-A-R Plug by any lab participating in each of five interlaboratory studies.



(a)



(b)

FIG. 12. Ranges of the means (a) and the standard deviations (b) of attenuation for the E-A-R Plug in five interlaboratory studies.

frequencies. The ranges of standard deviations for the S12/WG11 data are below the maximum range at all test frequencies and below the average range at five of the seven frequencies.

B. V-51R earplug data

Presented in Fig. 13 are the minimum and maximum mean attenuations from labs participating in three studies which included the V-51R earplug. The EPA and AF survey data are very similar. The S12/WG11 data show less attenuation, as demonstrated by the lack of overlap with the other two studies. The interlab variability of the V-51R data from these three sources is seen in the ranges of the means and standard deviations of attenuation shown in Fig. 14. The ranges of the mean attenuations increase with frequency above 250 Hz to values as high as 12 dB. At 250–1000 Hz the differences among studies are smallest. At the other frequencies, where differences among studies are larger, the S12/WG11 ranges are smaller than the other studies except at 2000 Hz. The ranges of all the standard deviations from the interlaboratory studies are less than 5 dB except for the EPA data at 8000 Hz. The ranges of standard deviation values for the S12/WG11 SF data are lowest at all test frequencies except 1000 and 2000 Hz.

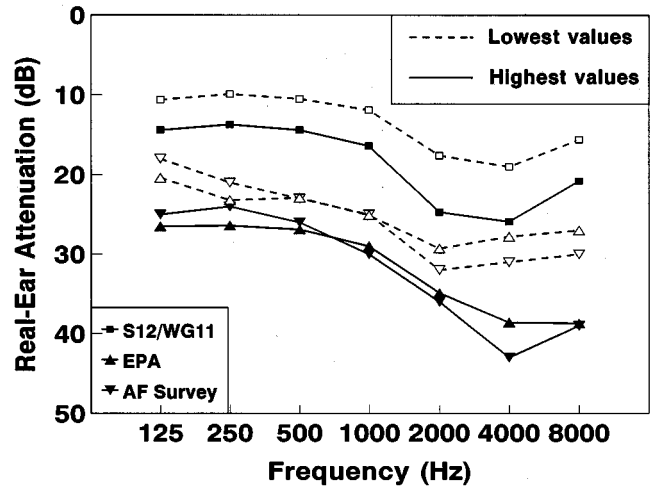
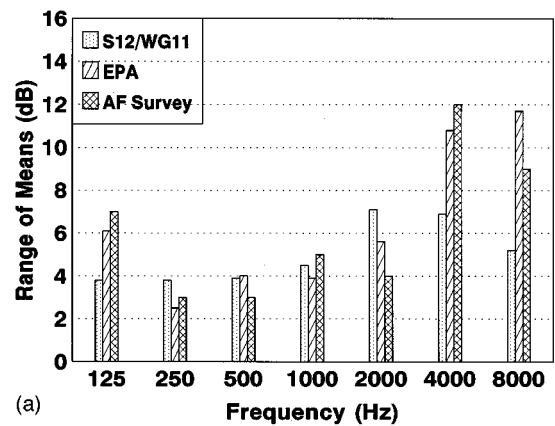


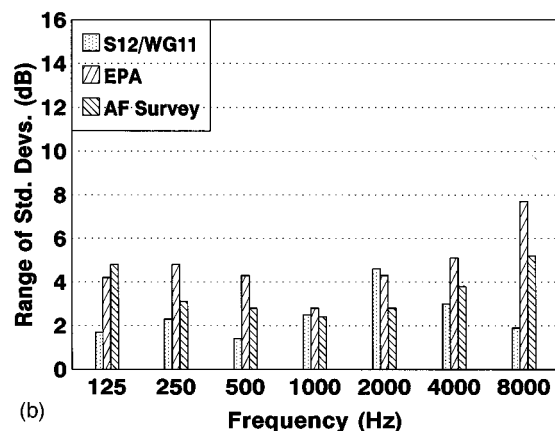
FIG. 13. Lowest and highest mean attenuation values obtained for the V-51R earplug by any lab participating in each of three interlaboratory studies.

C. Earmuff data

Data for four different earmuffs were examined from five different studies. One earmuff, the Optigard, was tested in two studies, but all others were tested in one study each. The lowest and highest single-lab mean attenuation values from each study are shown in Fig. 15. The lowest mean attenuation values from different studies are similar, except



(a)



(b)

FIG. 14. Ranges of the means (a) and the standard deviations (b) of attenuation for the V-51R earplug in three interlaboratory studies.

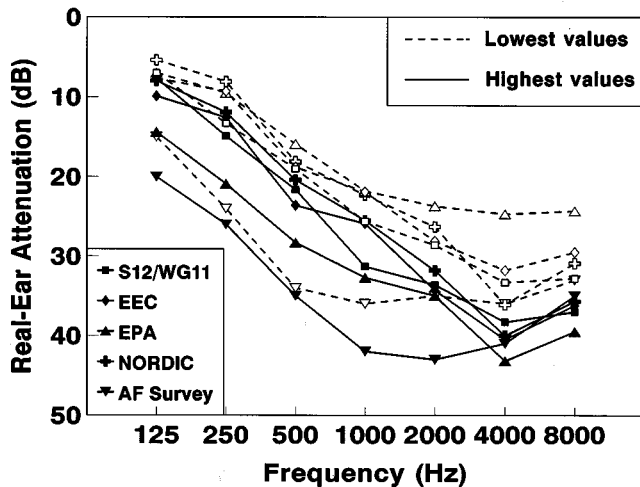


FIG. 15. Lowest and highest mean attenuation values obtained for the UF-1 earmuff by any lab participating in each of five interlaboratory studies.

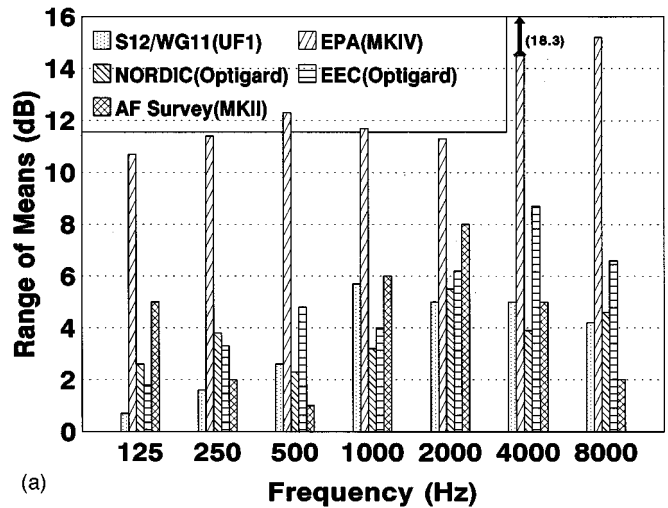
for the AF survey results, which indicate higher attenuation through 2000 Hz. The ranges of the means and standard deviations of attenuation for the four different earmuffs are shown in Fig. 16. No standard deviation data were provided for this device by the EEC study. The EPA study clearly displays the highest interlab variability, possibly due to the under-the-chin positioning of the earmuffs (instead of over-the-head). The ranges of mean attenuations from the remaining studies are similar, with no consistent rank-ordering. The S12/WG11 study displayed the smallest range of mean attenuation values at 125, 250, and 2000 Hz and never displayed the highest range values at any frequency. The S12/WG11 standard deviation ranges are lowest at 500, 4000, and 8000 Hz.

V. DISCUSSION

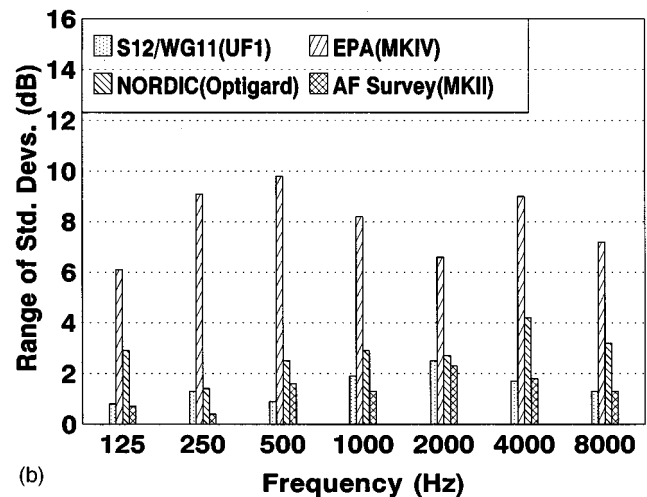
The S12/WG11 interlaboratory study provided a wealth of data, some of which have been examined in this paper. The results have immediate implications for the development of a new HPD test standard, as well as practical ramifications for the hearing conservation community.

A. Suitability of the subject fit protocol for standardization

The goal of S12/WG11's interlaboratory study was to assess whether either the SF or IUF procedure would yield useful estimates of achievable field attenuation for HPDs. Based on the pilot study, S12/WG11 anticipated that the SF protocol would yield mean attenuations of the desired magnitude, but there was concern that interlaboratory variability might be excessive, especially if subjects naive in the use of hearing protection were used. This concern proved unfounded, as results of the full-scale study showed that interlaboratory variability for the SF protocol was no higher than for the IUF protocol, and no higher than that achieved using other HPD test standards with subjects who had potentially greater experience in HPD use. The mean attenuations measured in the SF condition also corresponded more closely to real-world data than did the IUF measurements. A separate



(a)



(b)

FIG. 16. Ranges of the means (a) and the standard deviations (b) of attenuation for the UF-1 earmuff in three interlaboratory studies.

paper (Franks *et al.*, 1996b, in preparation) includes a detailed analysis of the degree of correspondence achieved between SF data and real-world results.

Based on the interlaboratory study, S12/WG11 is incorporating the SF protocol into a proposed new standard method for HPD attenuation testing. The use of subjects who are naive with respect to HPD use is a critical element of the SF protocol, since subjects' experience in HPD use strongly affects the attenuation they will achieve in an REAT evaluation (Berger, 1992).

B. Experience with the subject-fit protocol

Subjects' behavior in the SF condition demonstrated the woeful inadequacy of the instructions provided by some HPD manufacturers. When naive subjects were given the instructions and paid to read them, many did not comprehend how to use the devices properly. Each lab had at least one subject who inserted the EP100 plug backwards (handle-first into the earcanal), and seven other subjects (across all 96 subjects in four labs) asked the experimenter which end to insert (a question which the experimenter was not allowed to answer). Many subjects failed to notice that the small tab on the handle of the V-51R plug should be oriented in a certain

direction. Even after the demonstration by the experimenter in the IUF condition this subtle positioning factor was not evident to all subjects. For the E-A-R Plug, a few subjects did not compress or insert the plug correctly. For example, one subject rolled the foam plug into a ball (rather than a cylinder) and placed it in the concha rather than down into the ear canal. For earmuffs, some subjects placed the muff cushion on top of the earlobe rather than around it, while others folded their pinnae under the cushions. These experiences emphasize how important it is for manufacturers to design easy-to-use products and to provide clear and explicit fitting instructions and illustrations of correct HPD placement with important features emphasized. However, written instructions will never be the complete answer: In the real world, many HPD wearers will never glance at instructions, much less study them. Moreover, they may imitate incorrect techniques which they see used by other employees. Therefore, personal instruction in HPD use is vital if real-world hearing conservation programs are to be effective.

When subjects are not rejected for incompatibility with a certain product, design limitations of various HPDs are reflected in the attenuation results. For example, the V-51R plug must be inserted into the ear canal by pushing with a fingertip. If the subject has long fingernails, the plug may stick to the nail and be partially withdrawn along with the finger. In a similar fashion, creation of a vacuum between the finger and the plug may cause it to pull out when the finger is withdrawn. The EP100 plug's size designated "large" was not large enough for some subjects in this study (a problem which may also occur for other earplug designs). One subject was unwilling to insert any earplugs very deeply because plugs triggered a cough reflex. Protocols with larger numbers of unselected subjects increase the probability that mismatches will affect the attenuation results. Incompatibilities of this nature are common in real hearing conservation programs. Although ideally a different device should be offered to HPD wearers who experience such problems, this happens only in programs where HPD fit is monitored closely enough to identify a mismatch.

C. HPD performance as a function of fitting instruction

The improvement in attenuation for the IUF condition over the SF condition documents what many hearing conservationists already know: People receive much more protection from HPDs if they are individually fitted and taught how to wear their devices properly. Clearly, a major reason for poor HPD performance in the real world is that many noise-exposed personnel do not receive personal instruction in HPD use. Corroborating results found previously (Casali and Epps, 1986), the current study quantifies the increase in attenuation that might result if individual personal instruction were provided to HPD wearers.

The IUF condition was not intended to represent the ideal fitting and teaching situation, since the experimenter could not touch the subjects to assess fit or guide subjects' HPD placement, and could not coach subjects with tips not included in manufacturers' instructions. These limitations were placed upon experimenters to reduce interlaboratory

variability. In the opinion of WG members, differences in experimenter judgement and experience with particular devices, and resulting differences in their interactions with subjects, comprise one of the largest sources of interlaboratory variability in current test standards. In this study the SF condition eliminated this effect, while the IUF condition allowed some influence (through the experimenter's modeling of HPD use). Although the IUF condition yielded higher attenuation than the SF condition, the interlaboratory variability increased as well. It is the *interlaboratory* variability, not the *intralaboratory* variability, that is the critical parameter affecting reproducibility of test data (Berger, 1988). Because HPD manufacturers may choose to have their products tested in different labs, the ideal test protocol would exhibit minimal interlab variability. The results of this study indicate that the SF protocol does not increase interlab differences compared to the IUF protocol or to other test procedures in existing test standards.

VI. CONCLUSIONS

Two new protocols for hearing protector attenuation testing were defined and compared using three types of earplugs and one type of earmuff in a repeated-measures interlaboratory study in which four labs participated. The subject fit (SF) protocol was designed to minimize experimenter influence in order to assess the protection achieved by audiometrically experienced subjects who were naive with respect to using HPDs, when following the instructions provided by HPD manufacturers. The informed user fit (IUF) protocol allowed the experimenter to demonstrate HPD use and coach the subjects toward fitting the device consistent with manufacturers' instructions, while not actually touching the subject. The major findings of the investigation were the following.

(1) For earplugs, subjects achieved significantly higher attenuation in the IUF fitting instruction condition compared to the SF condition, but for earmuffs there was no practically important difference in attenuation between conditions.

(2) The interlaboratory variability was smaller in the SF condition than in the IUF condition, with a smaller range of mean attenuations across labs in SF and equivalent ranges of within-lab standard deviations in both conditions.

(3) Limited practice within the same fitting condition *without additional instruction* in fitting the HPDs did not significantly increase the attenuation achieved.

(4) There was no effect of the order in which subjects tested different devices.

(5) Wearing safety glasses together with earmuffs reduced attenuation by about 5 dB at all frequencies.

(6) Although the measurement of earmuff band force is an objective procedure implemented on a test fixture, the variation in measurements of that value across laboratories is on the order of 15% or more of the mean measured values.

(7) When the interlaboratory variability of data from the SF condition of this study was compared to that seen in four previous interlaboratory studies employing different protocols, the SF ranges of means and standard deviations of attenuations compared favorably to past studies. The SF ranges of standard deviations were smallest at most test frequencies

for the three types of devices common to all studies. The variability of the SF procedure as implemented in this study is no higher than, and sometimes less than, that observed among the data collected with other fitting procedures in other interlaboratory studies.

VII. NEXT STEPS FOR S12/WG11

Initial comparisons of the SF results from this investigation to the available real-world attenuation data for the same HPDs indicate that the SF protocol did yield attenuation values which correspond to the upper range of real-world findings (Franks *et al.*, 1996a, in preparation). Therefore, S12/WG11 is drafting a proposed new standard using the SF protocol to estimate achievable real-world attenuation.

Challenges to be met if the SF procedure becomes accepted include difficulty in recruiting subjects naive in the use of hearing protection; the ongoing dependence on manufacturers' instructions which may be incomplete or unclear; and the political, economic, and legal ramifications involved in changing government policies and regulations to make more accurate estimates of real-world attenuation available to HPD purchasers and users.

ACKNOWLEDGMENTS

This study would not have been possible without the careful, patient, and dedicated efforts of those individuals who collected and recorded the data (a process requiring over 320 h in each lab) and maintained cheerful rapport with the subjects through the process. Each of these individuals is gratefully acknowledged for this considerable effort: at Armstrong Laboratory, U.S. Air Force personnel Captain Nancy Green and Captain Denise West; at NIOSH Taft Laboratories, Curtis W. Sizemore; and at U.S. Army Aeromedical Research Laboratory, Barbara A. Murphy. Edward F. Krieg of NIOSH is acknowledged for his assistance with statistical analysis of the data.

ANSI (1974). ANSI S3.19-1974 (ASA STD 1-1975), "Method for the Measurement of Real-Ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs" (Acoustical Society of America, New York).

ANSI (1984). ANSI S12.6-1984, "Method for the Measurement of the Real-Ear Attenuation of Hearing Protectors" (American National Standards Institute, New York).

Berger, E. H. (1986). "Review and Tutorial—Methods of Measuring the Attenuation of Hearing Protection Devices," *J. Acoust. Soc. Am.* **79**, 1655–1687.

Berger, E. H. (1988). "Can Real-World Hearing Protector Attenuation be Estimated Using Laboratory Data?," *Sound Vib.* **22**(12), 26–31.

Berger, E. H. (1990). "Procedure for an Interlaboratory Comparison of REAT Protocols Intended to Provide an Improved Estimate of the Field Attenuation of Hearing Protection Devices," Cabot Safety Corp. Tech. Rept. E-A-R 89-14/HP, Indianapolis, IN.

Berger, E. H. (1992). "Development of a Laboratory Procedure for Estimation of the Field Performance of Hearing Protectors," in *Proceedings, Hearing Conservation Conference* (Off. Eng. Serv., Univ. Kentucky, Lexington, KY), pp. 41–45.

Berger, E. H. (1993a). "EARLog #20—The Naked Truth About NRRs," Cabot Safety Corp., Southbridge, MA.

Berger, E. H. (1993b). "Development of a new hearing protector test standard—Overview of the efforts of ANSI S12/WG11," *J. Acoust. Soc. Am.* **94**, 1791(A).

Berger, E. H., Franks, J. R., and Lindgren, F. (1996). "International review of field studies of hearing protector attenuation," in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Axelsson, H. Borchgrevink, R.

P. Hamernik, L. Hellstrom, D. Henderson, and R. J. Salvi (Thieme, New York), Chap. 29, pp. 361–377.

Berger, E. H., Kerivan, J. E., and Mintz, F. (1982). "Interlaboratory Variability in the Measurement of Hearing Protector Attenuation," *Sound Vib.* **16**(1), 14–19.

Berger, E. H., and Rowland, L. D. (1989). "Is it Necessary to Measure Hearing Protector Attenuation at 3.15 and 6.3 kHz?," *J. Acoust. Soc. Am.* **85**, 1590–1594.

Berger, E. H., Royster, J. D., Casali, J. G., Merry, C. J., Mozo, B. T., and Royster, L. H. (1990). "Results of a Pilot Interlaboratory Comparison of REAT Measurements of Hearing Protectors, Coordinated by ANSI S12/WG11," *J. Acoust. Soc. Am. Suppl.* **1** **88**, S11.

Casali, J. G., and Epps, B. W. (1986). "Effects of User Insertion/Donning Instructions on Noise Attenuation of Aural Insert Hearing Protectors," *Hum. Factors* **28**(2), 195–210.

Casali, J. G., and Park, M. Y. (1991). "Laboratory versus Field Attenuation of Selected Hearing Protectors," *Sound Vib.* **25**(10), 28–38.

Dixon-Ernst, C., and Behar, A. (1993). "Why users need accurate real-world estimates of hearing protectors' ratings," *J. Acoust. Soc. Am.* **94**, 1791(A).

Environmental Protection Agency (1979). "Noise Labeling Requirements for Hearing Protectors," *Federal Register*, Vol. 42, No. 190, 40 CFR Part 211, 56139–56147.

Franks, J. R., Berger, E. H., Krieg, E. F., Behar, A., Casali, J. G., Dixon-Ernst, C., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (1996a). "Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices, Part II: Sample size necessary to provide acceptable reproducibility," manuscript in preparation for submission to *J. Acoust. Soc. Am.*

Franks, J. R., Berger, E. H., Behar, A., Casali, J. G., Dixon-Ernst, C., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (1996b). "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," manuscript in preparation for submission to *J. Acoust. Soc. Am.*

Franks, J. R. and Casali, J. G. (1993). "Hearing protector attenuation from subject-fit methods at the work site and in the laboratory," *J. Acoust. Soc. Am.* **94**, 1791(A).

Franks, J. R. and Merry, C. J. (1993). "Sample size necessary to provide acceptable reproducibility in laboratory hearing protector attenuation testing," *J. Acoust. Soc. Am.*, **94**, 1792(A).

Howell, K. and Martin, A. M. (1973). "Prediction of experimental design for the standard measurement of the acoustic attenuation of hearing protectors," *Inst. of Sound and Vibration Res.*, Memo No. 472, Southampton, England.

ISO (1981). ISO 4869:1981, "Measurement of Sound Attenuation of Hearing Protectors" (International Organization for Standardization, Geneva, Switzerland).

ISO (1990). ISO 4869-1:1990(E), "Acoustics—Hearing Protectors—Part 1: Subjective Method for the Measurement of Sound Attenuation" (International Organization for Standardization, Geneva, Switzerland).

Johnson, D. L. (1976). Unpublished letter of 9/23/76 to V. Tetrolati of the EPA Office of Noise Abatement and Control in response to request for information.

Merry, C. J., Sizemore, C. W., and Franks, J. R. (1992). "The effect of fitting procedure on hearing protector attenuation," *Ear Hear.* **13**, 11–18.

Nixon, C. W. (1993). "Comparison of WG11 study variability to that of other interlaboratory studies," *J. Acoust. Soc. Am.* **94**, 1792(A).

Occupational Safety and Health Administration (1983). OSHA Instruction CPL2-2.35A, December 19, 1983 (update to the Field Operations Manual). Government Institutes, Inc., Rockville, MD.

Ohlin, D. (1993). "A discussion of WG11's anthropometric data with comparison to other available databases," *J. Acoust. Soc. Am.* **94**, 1792(A).

Poulsen, T. (1984). "Nordic Round Robin Test on Hearing Protector Measurements, Vol 1, Subjective Results," Internal Report No. 21 1984, Acoustics Laboratory, Technical University of Denmark, Lyngby.

Royster, J. D. and Royster, L. H. (1993). "Results of the S12 Working Group 11 interlaboratory study to approximate real-world attenuation of hearing protection devices (HPDs)," *J. Acoust. Soc. Am.* **94**, 1791(A).

Shipton, M. S. (1986). "Intercomparison of Measurements on Ear Protectors by Subjective and Objective Methods," Contract No. 778/1/000/057/82/4-BCR-UK(30), Commission on European Communities, Luxembourg.

Winer, B. J. (1971). *Statistical Principles in Experimental Design* (McGraw-Hill, New York).