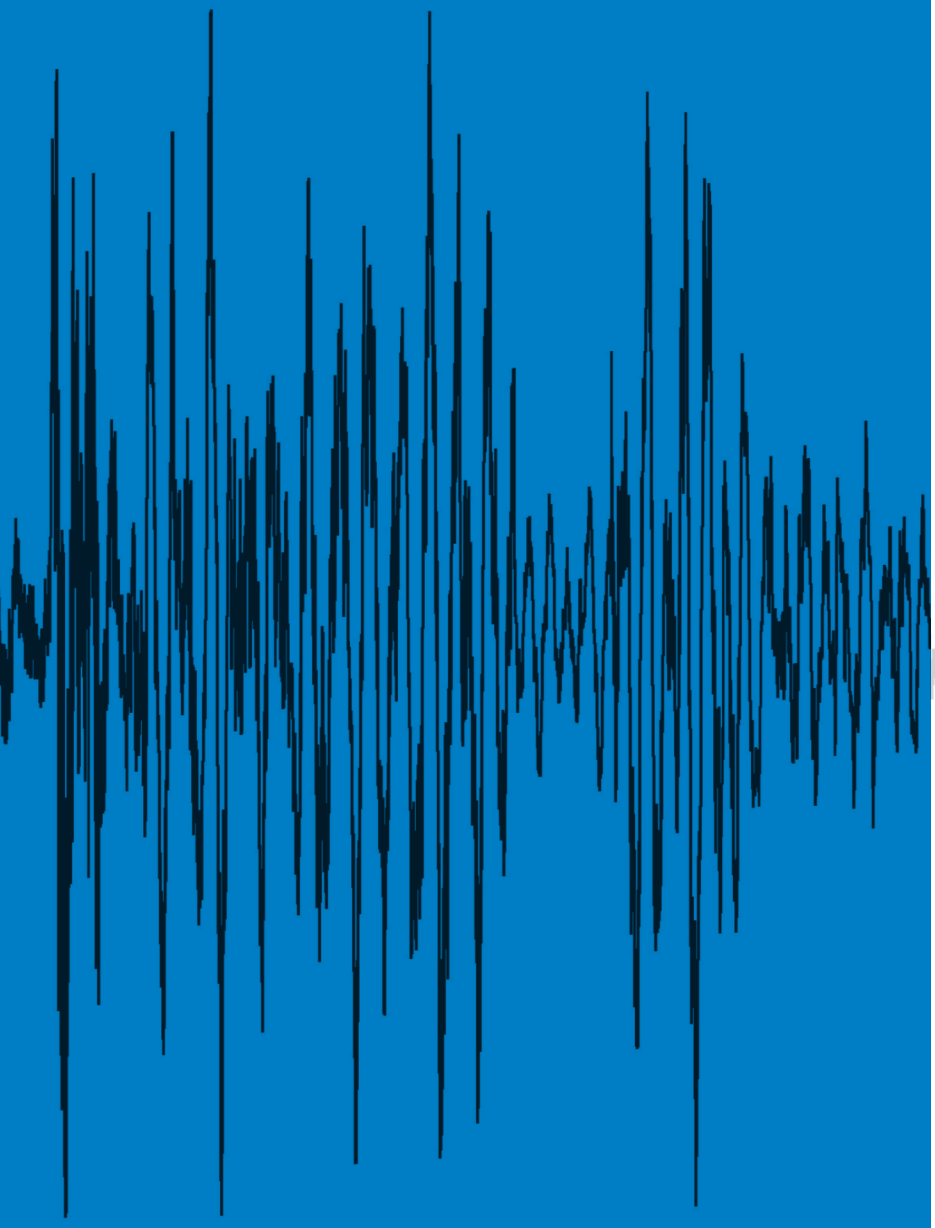
87,091

ISSN 1463-1741

Indexed with MEDLINE, EMBASE & SCI



Noise & Health

A Bi-monthly Inter-disciplinary International Journal

www.noiseandhealth.org

March-April 2011 | Volume 13 | Issue 51

Medknow Publications

Development and validation of a field microphone-in-real-ear approach for measuring hearing protector attenuation

E. H. Berger, J. Voix [1], R. W. Kieper, C. Le Cocq[1]

3M Occupational Health and Environmental Safety Division, E•A•RCAL Laboratory, 7911 Zionsville Road, Indianapolis, IN 46268-1657, [1]École de Technologie Supérieure, Department of Mechanical Engineering, 1100 Notre-Dame West, Montreal, H3C 1K3, Canada

Abstract

Numerous studies have shown that the reliability of using laboratory measurements to predict individual or even group hearing protector attenuation for occupationally exposed workers is quite poor. This makes it difficult to properly assign hearing protectors when one wishes to closely match attenuation to actual exposure. An alternative is the use of field-measurement methods, a number of which have been proposed and are beginning to be implemented. We examine one of those methods, namely the field microphone-in-real-ear (F-MIRE) approach in which a dual-element microphone probe is used to measure noise reduction by quickly sampling the difference in noise levels outside and under an earplug, with appropriate adjustments to predict real-ear attenuation at threshold (REAT). We report on experiments that validate the ability of one commercially available F-MIRE device to predict the REAT of an earplug fitted identically for two tests. Results are reported on a representative roll-down foam earplug, stemmed-style pod plug, and pre-molded earplug, demonstrating that the 95% confidence level of the Personal Attenuation Rating (PAR) as a function of the number of fits varies from ±4.4 dB to ±6.3 dB, depending on the plug type, which can be reduced to ±3.1 dB to ±4.5 dB with a single repeat measurement. The added measurement improves precision substantially. However, the largest portion of the error is due to the user's fitting variability and not the uncertainty of the measurement system. Further we evaluated the inherent uncertainty of F-MIRE vs. the putative "gold standard" REAT procedures finding, that F-MIRE measurement uncertainty is less than one-half that of REAT at most test frequencies. An An American National Standards Institute (ANSI) working group (S12/WG11) is currently involved in developing methods similar to those in this paper so that procedures for evaluating and reporting uncertainty on all types of field attenuation measurement systems can be standardized. We conclude that the hearing conservationist now has available a portable, convenient, quick, and easy-to-use system that can improve training and motivation of employees, assign hearing protection devices based on noise exposures, and address other management and compliance issues.

Keywords: Hearing protection, fit testing, field attenuation, real-world attenuation, field performance

Introduction

When properly and consistently worn, hearing protectors can effectively attenuate noise and prevent hearing. That much is clear. However, the devil is in the details – how to train employees to properly and consistently wear their hearing protection devices (HPDs), how to suitably assign HPDs commensurate with noise exposures, and how to accommodate personal preferences and anatomical considerations. Success in hearing protection fitting and use takes care and awareness to detail as well as individualized attention. Heretofore, this was complicated by the fact that those dispensing hearing protection in industry had little or no training in how to fit the HPD[1] and that the only noise attenuation data available were from group-average data based on laboratory measurements, as reflected in the Noise Reduction Rating (NRR). Even if the laboratory data were representative of the actual group of subjects using the device, individual variability is large enough that attempts at predicting one person's performance from group data can easily err by up to 20 dB.[2]

One approach to solving these problems is the development of systems to allow individual fit testing in industry, and indeed such systems have proliferated. Fit-test technology has been available in the laboratory in many forms for nearly 30 years. Berger began publishing in this realm in 1984,[3-6] but only in the past decade has the wider hearing conservation community started to look more closely at this issue.

Access this article online	
Quick Response Code:	Website: www.noiseandhealth.org
	DOI: 10.4103/1463-1741.77214

Recently, Berger discussed seven important applications for field-test methods, as listed below.[7]

1. Train and motivate employees to properly and consistently wear their HPDs.
2. Train the person responsible for fitting and training employees on how that should be done.
3. Assign HPDs based on noise exposures and expected protection levels.
4. Provide a useful standard-threshold-shift (STS) follow-up procedure that could be used to determine whether the problem may be HPD related.
5. Provide data that may be accepted by the U. S. Occupational Safety and Health Administration (OSHA) as a better alternative to using labeled attenuation values and derated NRRs to assess HPD adequacy.
6. Audit departments to evaluate overall HPD effectiveness and suitability.
7. Provide potentially useful documentation to help defend against workers' compensation claims alleging inadequate hearing protection and insufficient training.

Today, there are a number of systems that provide field-test capabilities. Although such systems do provide enhanced prediction of individual protection performance, they include their own errors and uncertainties inherent to the measurement process and user-fitting capabilities. The purpose of this paper is to explore and characterize the accuracy and precision of one method, microphone-in-real-ear (MIRE), and its implementation as a quick and portable field method, termed field-MIRE, abbreviated F-MIRE.[8] We will consider how to validate and qualify an F-MIRE system and how to provide an appropriate means of recognizing and addressing the inherent variability that is still present, even in field-test methods, and how this may be addressed in a planned ANSI standard on field attenuation measurement systems (FAMS).[9]

MIRE is an objective approach to measuring hearing protector attenuation that has been used in the laboratory for many years,[4] is described in an ANSI standard,[10] and more recently has been applied in the field as F-MIRE. With F-MIRE, the sound pressure levels in the earcanal under the hearing protector as well as those outside the HPD are simultaneously measured. Using suitable correction factors (see next section) to account for known and quantifiable acoustic differences between the F-MIRE and the standardized real-ear attenuation at threshold (REAT) data, the objectively measured values can be used to accurately estimate the hearing protector's attenuation.

One embodiment of the F-MIRE approach is 3M's E•A•Rfit™ Validation System, which is evaluated in this report. It incorporates a small dual-element microphone and the associated patent-pending proprietary technology.[11-13] One section of the dual-element microphone is coupled to a tube that passes through the earplug to pickup the sound pressure levels in the earcanal and the other section measures the external sound field. Using a special probed earplug (a "surrogate" earplug), it only takes about 10 s to take the measurement for any one fit in one ear at the seven standard test frequencies from 125 Hz to 8 kHz and to calculate the associated NRR, referred to as the Personal Attenuation Rating (PAR).

The PAR is computed like the Noise Reduction Statistic for use with A-weighting (NRS_A), which is defined in ANSI S12.68-2007,[14] with the exception that the between-subject variability is replaced by the sum of the variances of the F-MIRE uncertainty and the within-subject re-fitting uncertainty. Thus, Equation (6) in ANSI S12.68 is replaced by Equation (1) below, wherein x is selected appropriately for the 80th and 20th desired percentile, respectively, $\overline{\text{ATT}}$ is the average corrected F-MIRE value across fits for a given subject (i.e., predicted REAT), and α is 0 or ± 0.84, depending on whether the mean or 80th or 20th percentile is selected. The variable, s_{spectrum} as is described in the ANSI standard, is defined using Equation (2), where N_n is the total number of the 100 noise spectra specified in that standard and n_n is the index of the spectrum used. The F-MIRE prediction uncertainty ($s^2_{\text{F-MIRE}}$), Equation (3) (also see Voix and Hager[15]) represents the difference between the F-MIRE and the REAT values averaged across the prior N_p (subjects x fits) measurements with laboratory subjects, that were conducted to establish the compensation factors as described in the next section. Fitting uncertainty is given by Equation (4) below. In Equation (4), m_f is the measured F-MIRE for each fit of the earplug, denoted by index n_f, and $\overline{\text{ATT}}$ is the average F-MIRE across all N_f fits and all N_n spectra for that subject, as was defined for Equation (1).

$$\text{PAR}_x = \overline{\text{ATT}} - \alpha_x \sqrt{s^2_{\text{F-MIRE}} + s^2_{\text{fit}} + s^2_{\text{spectrum}}} \quad \text{....(1)}$$

$$s_{\text{spectrum}} = \sqrt{\frac{1}{N_n - 1} \sum_{n_n=1}^{N_n} \left(m_n - \overline{\text{ATT}}\right)^2} \quad \text{....(2)}$$

$$s_{\text{F-MIRE}} = \sqrt{\frac{1}{N_p - 1} \sum_{n_p=1}^{N_p} \left(\text{ATT}_{n_p} - \text{REAT}_{n_p}\right)^2} \quad \text{....(3)}$$

$$s_{\text{fit}} = \sqrt{\frac{1}{N_f - 1} \sum_{n_f=1}^{N_f} (m_f - \overline{\text{ATT}})^2} \quad \text{....(4)}$$

Even though F-MIRE prediction uncertainty, as shown in Equation (3), is much less than the inherent uncertainty in the classical approach of using mean laboratory data on one group of subjects to make field predictions for a given individual (who was not part of that group), it is important that it be reported and understood. The amount of variability in PAR is in part the subject of this paper.

Figure 1 illustrates the components of the F-MIRE system used in this research and Figure 2 depicts the microphone and probed earplug tips. The F-MIRE system consists of a sound source that can generate high levels of broadband random

noise at the listener's ear, a dual-element microphone that simultaneously measures in a repeatable location the sound present at the outside of the earplug and that present in the earcanal after having passed through the earplug, a probed earplug to act as a surrogate for the actual earplug that subjects actually wear in use, and an analysis system installed onboard a digital signal processing board inside the loudspeaker enclosure. The speaker/analysis system, which is connected to a desktop or laptop PC, takes measurements in typically <10 s for each fit. The sound levels used, depending on the level of attenuation provided by the earplug, range from about 85 dBA to 95 dBA. The listener's nose is positioned 30 cm from the front of the loudspeaker.

A key feature of the development of this F-MIRE system is the design of the probed test tips. The tubing through the plug must allow measurement of the sound pressure levels in the earcanal via the dual-element mic but must, at the same time, have high levels of self-insertion loss (i.e., sound transmission through the wall of the tubing as opposed to sound transmission through the lumen of the tubing). The tubing must also be of sufficiently small diameter and adequate softness that it does not materially affect the listener's ability to insert the earplugs, yet with an adequately large inside diameter so that it does not have too high an inertance that would excessively attenuate the transmission of high-frequency sound. In the case of the foam tips, the tubing also must not detract from the ability to roll the plug into a tiny crease-free cylinder for insertion into the earcanal.

Procedures

F-MIRE measurements yield what is termed noise reduction (NR), which is the difference between the levels measured outside and inside the earcanal. REAT, on the other hand, can be considered a subjective insertion loss (IL) measurement that indicates the difference in the hearing threshold levels at one point in space (namely the eardrum) with and without the HPD in place. NR and IL are directly related, but they are not the same, more so because of the plane of measurement issues than because of the type of measurement; thus a mathematical adjustment is required that uses the transfer function of the open ear (TFOE), which is the difference between the sound pressure levels in the sound field and at the eardrum. See Berger[4] for details. Besides the TFOE correction, sound conduction through the small lumen of the probe tube tips varies with frequency, and this must also be taken into account. Other correction factors to account for the occluded earcanal resonance and the absence of bone-conduction transmission and physiological noise masking (that are present in REAT measurements) are also necessary.[13]

The most direct way to account for all the above factors is to make a simultaneous measurement of REAT and NR for a given fitting of the probed earplugs on a group of subjects. One can then directly compare the measured values of attenuation and determine the best correction (also called compensation) factors to bring them into the closest possible agreement.[11] This type of approach is commonly accepted and has previously been used for other types of field test systems.[16]

The compensation factors noted above only describe the differences between REAT and NR due to system bias, factors that are stable from measure to measure. There is also the question of the variability of the measurement systems and how those may differ. Accounting for this multiplicity of factors requires separate experiments as discussed below.

Test facility and equipment

REAT tests were conducted in a 113 m^3 reverberant chamber with procedures in accordance with ANSI S12.6-2008,[17] with the exception of the method of fitting hearing protectors, which is described later in this section. The facility is accredited under the Department of Commerce, National Voluntary Laboratory Accreditation Program for testing to the ANSI standard.[18]

The E•A•RFit system hardware was described previously. The version of the software used in these laboratory experiments was initially 2.2.0, but for some of the later experiments version 3.2.1.5 was used, which included updated compensation factors. The F-MIRE measurements were conducted in the E•A•RCAL laboratory immediately outside the test chamber. Sound levels are not controlled in that relatively quiet space, but neither are they critical for purposes of F-MIRE testing with the E•A•RFit system. The background noise levels in the laboratory are representative or more likely lower than what might be encountered in a typical office or safety facility where field testing of HPDs occurs.

Test materials

In this series of experiments, the goal was to develop compensation factors and to evaluate the performance of the E•A•RFit system and the accompanying probed earplugs, also called "tips," that are provided specifically for use with that system. The work in this paper focuses on three of the tips that are representative of most of the styles that can be tested with this system. They are the Classic™ roll-down foam earplugs, the Push-Ins™ stemmed-style pod plugs, and the UltraFit™ pre-molded earplugs. The data are typical of those for the other types and styles of products that we have evaluated.

Test subjects

Test subjects and their interaction with the experimenter are key to the results obtained in laboratory HPD attenuation measurements.[19] Although the goal here was to obtain predictive results for field application of F-MIRE, we chose

to work with a trained panel of listeners. This was because our prior research indicated that inexperienced subjects appropriate for Method-B testing, as described in the ANSI 12.6-2008 REAT standard,[17] are not reliable enough in general to provide consistent results for the extensive testing required for the experiments of this study. Moreover, with the amount of fitting, refitting, and controlled fitting necessary to get the desired levels of performance for development of predictive data over a wide range of attenuation values, inexperienced subjects would have quickly become experienced.

Experimental procedure

The experimental protocol is summarized in Table 1, which is separated into three sections to clarify the different aspects of the protocol. Experiment #1 was to measure the accuracy of the E•A•Rfit predictions and compensation factors, Experiment #2 to assess fitting variability, and Experiment #3 to compare the REAT and F-MIRE uncertainty. Twenty experienced subjects drawn from the E•A•RCAL pool of subjects participated in each of the experiments.

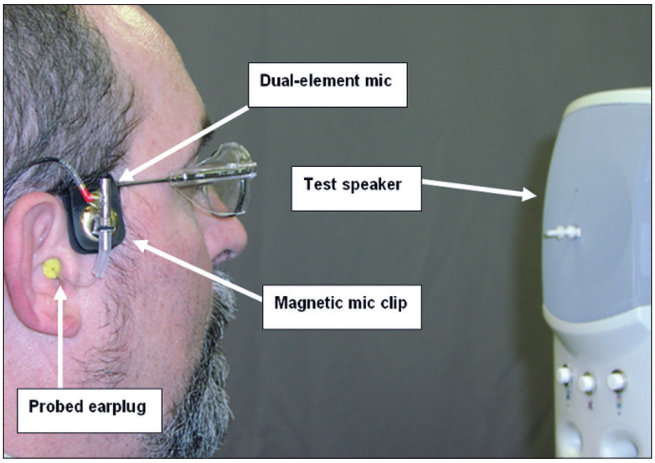


Figure 1: Key components of the F-MIRE system

Figure 3: Earplug with probe, with brass nipple inserted to seal the probe at the distal end

In Experiment #1, each subject entered the chamber and began with a REAT evaluation of a probed plug (an open followed by an occluded threshold). For the occluded test, the plug was fit by the experimenter and the opening in the probe tube at its distal end was sealed with a brass plug as shown in Figure 3. This provided a measurement of the attenuation of the probed plug that would reflect any flanking pathways through the walls of the tubing and the connecting sleeve. Thus, if the tube degraded the performance of the earplug itself, it would be apparent by comparing this measurement with that of an unmodified (i.e., standard) earplug as discussed below.

The fit was controlled by the experimenter because the goal in this experiment was to obtain two distinctly different fits of the plug for each subject in order to obtain a measure of the correspondence between the REAT and the F-MIRE data over a wide a range of attenuation values. As our goal was not related to evaluating a subject's ability to fit the product, the

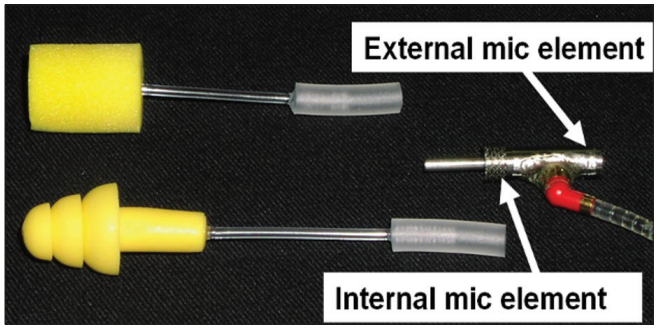


Figure 2: The dual-microphone element and representative probed tips for the foam and pre-molded earplugs

Figure 4: Examples of good (left), degraded (middle), and unacceptable (right) fits for Classic (top), Push-Ins (middle), and UltraFit (bottom) earplugs

Table 1: Outline of experimental plan to determine compensation factors and assess fitting variability

Experiment #1 – Accuracy: Experimenter fit, 2 repetitions of E1 through E8	
E1	REAT (open)
E2	REAT (occluded), good fit, probed plug sealed with brass nipple
E3	F-MIRE, same fit as E2, each ear one at a time, brass nipple replaced by F-MIRE microphone
E4	F-MIRE, degraded fit, each ear one at a time
E5	REAT, same degraded fit as above, probed plug sealed with brass nipple
E6	F-MIRE, same degraded fit as above, each ear one at a time
E7	REAT, open
E8	REAT, good fit, unmodified plug
Experiment #2 – Fit variability: Subject fit	
S1	F-MIRE, 5 measures, each ear, remove and refit plug and mic between each fit
Experiment #3 – Precision of F-MIRE: Experimenter fit	
P1	REAT (open), 3 repeat thresholds, everything held constant
P2	REAT (occluded), good fit, plug sealed with brass nipple, 3 thresholds, everything held constant
P3	F-MIRE, same fit as P2, plug remains in place but mic re-fitted for each of 3 repeat measures

fact that the experimenter inserted the plug was not an issue.

At the high end of performance, we obtained as good and as deep a fit as possible, similar to what would be achieved during a product labeling test per ANSI S3.19-1974[20] that would be in compliance with the current U.S. Environmental Protection Agency hearing protector labeling regulation.[21] At the low end of attenuation, the fitting for the degraded condition of fit discussed below, we purposely reduced the quality of the fit but did not totally corrupt the performance as the intent was to measure the poorest protection for which one might wish to use a fit-testing system. In cases where the plug's fit was so dreadful as to be visually apparent to all but the most unobservant or untrained fitter, it is unlikely that the use of a fit-testing device would be attempted. Even if it were, the positioning of the plug would be so unstable in the ear that it would be difficult to take a reliable or consistent reading to compare the REAT and the F-MIRE conditions. Photographs illustrating the fitting depths are reproduced in Figure 4.

Following the REAT, the subject exited the chamber, the F-MIRE microphone was plugged into the probed plug, and objective attenuation measurements were taken for each ear. This first set of "paired" measurements provided the comparison between REAT and F-MIRE for a well-fitted earplug.

The plug's performance was then degraded and repeated F-MIRE values recorded until the desired level of reduced attenuation at 125 Hz was achieved. This varied by product, approximately 10–25 dB for the Classic, 10–20 dB for the Push-Ins, and 10–15 dB for the UltraFit. Once a desired fit was obtained, the last F-MIRE value was retained for documentation but was not used in the subsequent analyses. The microphone was removed from the plug, being as careful as possible not to dislodge the less-than-ideally seated earplug, the brass plug was re-inserted and the subject then re-entered the chamber for an occluded test of the degraded fit. Immediately thereafter, the subject exited the chamber for an F-MIRE test of the same fitting.

The second F-MIRE test, post-REAT testing, was examined to make sure that the fit of the plug had not changed from the initial F-MIRE measurement. It was also the measurement selected for comparison with the REAT data. The reason for selecting the second F-MIRE test was that it would conform most closely to actual field experience, in that subjects would fit plugs to their ears and then use the F-MIRE to find out how they did. An additional justification was that we were more likely to affect the fit of the poorly seated plug during removal of the probe microphone before entering the room, when we had to be careful not to tug on the plug and degrade its fit further prior to the REAT evaluation, than during insertion of the probe microphone after exiting the room and prior to the final F-MIRE evaluation. After the second F-MIRE test, the subject re-entered the chamber for a REAT evaluation (open and occluded threshold) on a well-fitted, unmodified (i.e., standard) earplug. The purpose of the last test was to assess the amount by which the attenuation provided by the fully sealed probed plug might fall short of a standard plug by virtue of having placed a tube through the product. The entire sequence described thus far was repeated a second time. That concluded the experimenter fit portion of the test sequence shown in the upper part of Table 1.

Experiment #2 was accomplished immediately outside the test chamber in the E•A•RCAL laboratory [Table 1, row S1]. It consisted of repeat F-MIRE measurements to assess the reliability of a subject's insertion of the probed plug. The plug was inserted by the subject and fitted with the probe microphone by the experimenter in one ear. This process was repeated four additional times in that ear, meaning that the microphone was removed from the plug and the plug removed from the ear, and then the earplug and microphone were sequentially re-inserted. This was then repeated in the other ear, providing a total of $N_f = 5$ measurements in each ear [see Equation (4)]. Across 20 subjects, this yielded five repeat measurements on 40 ears, and so concluded the full series of tests on each subject.

Experiment #3 was a separate uncertainty experiment, conducted using only a single product, the Classic foam earplug, to assess the inherent measurement uncertainty in REAT and F-MIRE data. Twenty subjects (not exactly the same group, but drawn from the same pool as in the preceding experiment) also participated in this study in which the subject entered the chamber and took three consecutive sets of unoccluded thresholds at the seven test frequencies. The experimenter then inserted the probed

foam plug for a good fit and, with the plug remaining in place and sealed by the brass tip, the subject took three consecutive occluded thresholds. The subject then exited the chamber and the F-MIRE microphone was fitted to the plug. Three consecutive F-MIRE measurements were then taken with the plug remaining in place but with the microphone removed and re-fitted between measurements. This allowed us to separately capture the F-MIRE measurement variability uncontaminated by fitting variability due to the positioning and re-positioning of the earplug.

Results I-Ability to predict REAT from F-MIRE values

When comparing REAT and F-MIRE data from Experiment #1, the question arises as to how to relate the binaural REAT measurement to the left- and right-ear diotic data developed by an F-MIRE system. In the REAT process, the dominant ear will be the one perceiving the highest sound levels. Ear dominance at each frequency will be controlled by the ear that has the least attenuation, offset by any differences between the absolute thresholds in the two ears. This is mathematically developed and presented by Voix and Laville.[13] In spite of our use of the computation of an equivalent binaural F-MIRE value for comparison with binaural REAT, in practice, when such a system is used for testing the fit of HPDs and for training, the immediate feedback to the subject is generally provided one ear at a time.

Figures 5–7 present the REAT measurements on the standard unmodified earplugs compared with the probed versions of those same plugs. In each of the three figures, the results for the unprobed and probed plugs are compared. The two versions of each plug provide attenuation values that agree within 2 dB at most test frequencies, with the largest divergence being 2.5 dB at 250 Hz for the UltraFit plugs. The differences were compared, frequency by frequency, using a paired *t*-test, and were found to be not significantly different at $P < 0.05$, except at 250 Hz and 2000 Hz for the UltraFit and 2000 Hz for the Push-Ins. An overall test of significance for each product, the Hotelling T^2 a multivariate version of the *t*-test, found a significant difference at $P < 0.05$ for the UltraFit product.[22] Regardless of the statistical significance, the differences were small in magnitude and unlikely to substantially affect the ability of the probed version of the plugs to provide a valid indication of the performance of a similarly fitted unmodified earplug of the same design. Furthermore, because the probed version of the UltraFit showed slightly lower attenuation than the standard version, any errors in measurement would be in the conservative direction of underestimating the true performance of the unmodified earplug.

Figures 8–10 present the averaged trends for the REAT vs. the equivalent binaural F-MIRE results that establish the validity of using the compensation factors from this study to make predictions of the REAT data from a MIRE measurement. We have separated the data by well-fitted and degraded-fit insertions. The solid lines are the REAT values and the dashed lines are the F-MIRE values, with the compensation included. The agreement for both the degraded and the good fits is within 2.5 dB for both fits for all devices, except at 2000 and 8000 Hz, with differences up to 2.9 dB for the Push-Ins, and of 3–3.6 dB at 500, 2000, and 8000 Hz for the Classic.

Of greater interest than the average values in Figures 8–10 are the scatter plots for individual PAR values and the trends that are observed. These data are presented in Figure 11 for the overall PAR, the value that most users would employ to make decisions. The good-fit values are shown with red circles and the degraded-fit with blue crosses. Superimposed on the data is the x = y linear relationship, as well as lines at $\pm$10 dB, to indicate points for which the prediction from the F-MIRE values would be divergent from REAT by >10 dB. There are 40 red circles and 40 blue crosses in each chart of the 20 × 2 good and 20 × 2 bad fits, four measurements on each of the 20 subjects. The compensation factors in the F-MIRE data have been adjusted for best fit. Ideally, the agreement between F-MIRE and REAT, i.e. the prediction error (defined by the difference between the reported REAT value and the predicted F-MIRE attenuation), should be independent of the level of the attenuation, and this is generally the case for these data. Note that for each plug, all but two measurements (three in the case of the Classic) agree within 10 dB, meaning that 97–98% of the values fall within the specified range.

If both values are equal for a given test point, the marker is exactly aligned on the x = y line and the prediction error (defined by the difference between the reported REAT value and the predicted F-MIRE attenuation) is zero. The accuracy with which F-MIRE predicts REAT is the distribution of the points around the x = y line in Figure 11 (they should cluster uniformly around the line, with no indication of a bias showing over- or under-prediction). However, it is questionable to examine individual points where there is disagreement, such as the red circle above the +10 dB line in the lower left panel, and presume that this reflects an erroneous F-MIRE prediction. This is because there is uncertainty in both F-MIRE and REAT measurements as we discuss in Results III, where data are presented on the relative magnitude of those uncertainties. Furthermore, computing a regression line for the scatter plots is not useful because individual outliers, even with these large data sets, can substantially change the slope of the curve. Thus, we selected an equivalency test to compare the two procedures.[23]

Although the four prediction-error datasets collected (a first and second fit of both the good and degraded insertions) do come from the same 20 subjects and could not, at first sight, be considered as independent, they are in

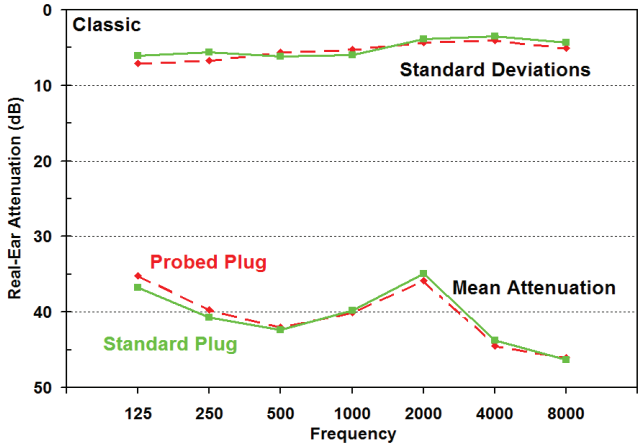


Figure 5: Mean REAT (lower part of chart) and standard deviation (upper part of chart) for the unmodified standard and probed Classic™ foam earplug

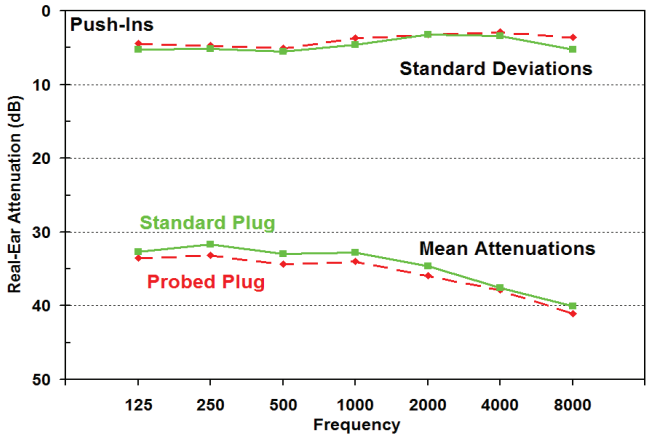


Figure 6: Mean REAT (lower part of chart) and standard deviation (upper part of chart) for the unmodified standard and probed Push-Ins™ earplug

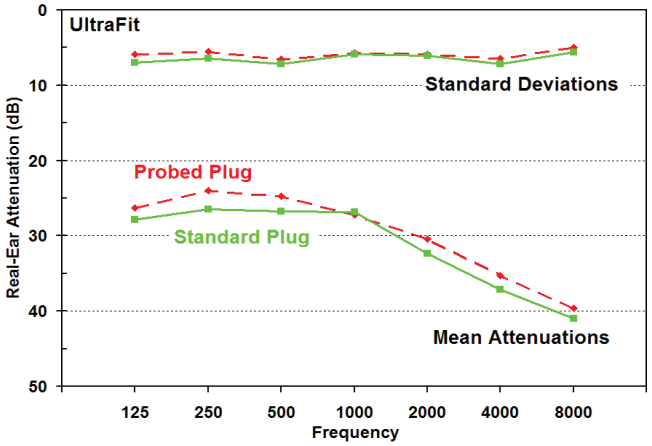


Figure 7: Mean REAT (lower part of chart) and standard deviation (upper part of chart) for the unmodified standard and probed UltraFit™ earplug

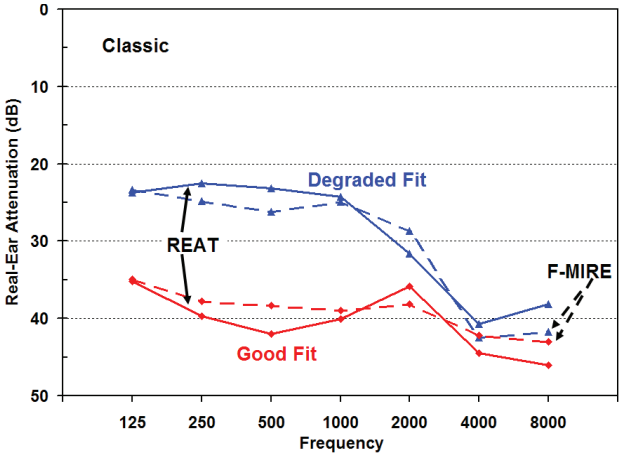


Figure 8: Comparison of corrected binaural F-MIRE predictions for the Classic™ foam earplug, using compensation factors determined in this study, with REAT data for the same fit for 20 subjects

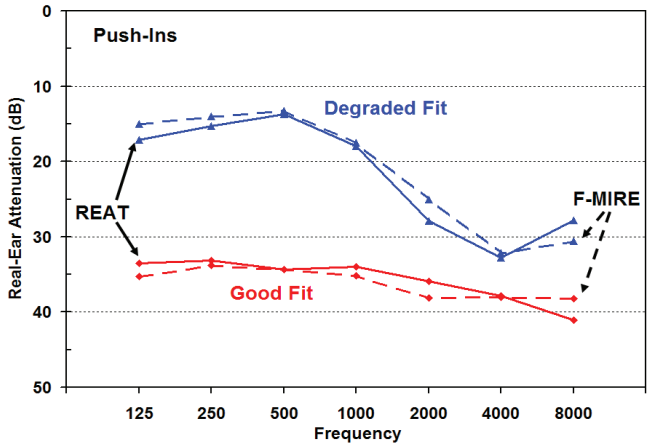


Figure 9: Comparison of corrected binaural F-MIRE predictions for the Push-Ins™ earplug, using compensation factors determined in this study, to REAT data for the same fit for 20 subjects

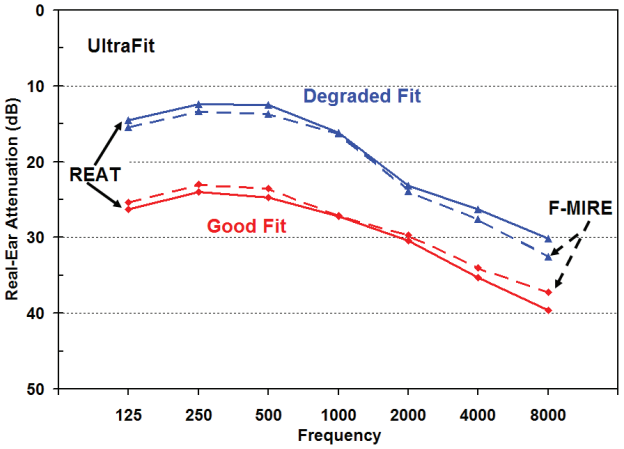


Figure 10: Comparison of the corrected binaural F-MIRE predictions for the UltraFit™ earplug, using compensation factors determined in this study, with REAT data for the same fit for 20 subjects

practice, however, only loosely pairwise correlated; deep-fit attenuation cannot linearly be related to shallow-fit attenuation, nor can first-fit attenuation predict second-fit attenuation. This can be explained by the fact that, in this experiment, the fitting was by the experimenter, not the subject, and thus the fit itself was the principal source of variance. We tested the four prediction-error distributions for normality, and each passed the Lilliefors normality test at a 5% significance level. Additionally, we compared all the first-fit data (for both good and degraded fits) to the second-fit data as well as all the good-fit data (for both first and second fits) to all the degraded-fit data, using the two-sample Kolmogorov-Smirnov goodness-of-fit test. At a 5% significance level, we found that in each of the four prediction-error distributions, the data were drawn from similar underlying populations. As a result, we computed the estimate of the mean and standard deviation of the prediction errors using the aggregate of the four datasets (80 data points). The standard deviations were 4.6, 3.6, and 4.8 dB respectively for the Classic, Push-Ins, and UltraFit products. These are the values of F-MIRE prediction uncertainty ($s_{F\text{-}MIRE}$) that may be used in Equation (1).

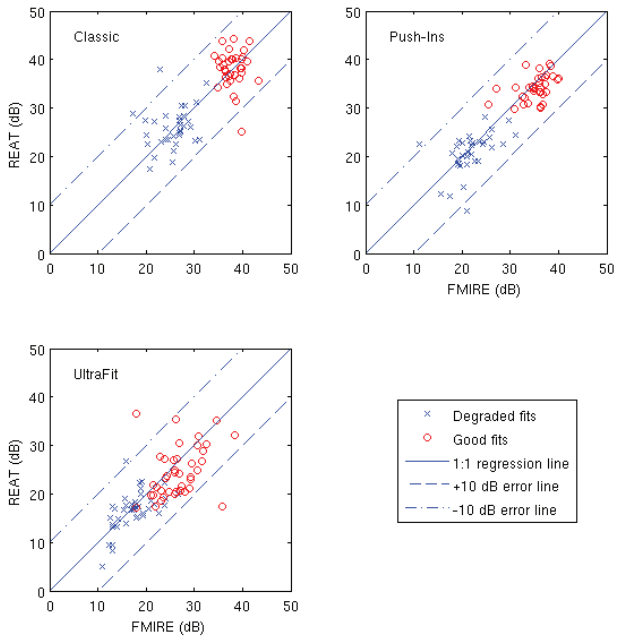


Figure 11: Scatter plots of REAT vs. binaural F-MIRE for the Classic™, Push-Ins™, and UltraFit™ earplugs for 20 subjects ×2 fits, with both a good and a degraded fit

To assess the differences between the REAT and the F-MIRE estimates, we established an interval of ±3 dB as the range within which the 95% confidence intervals for the parameter estimates on the mean prediction error should fall in order to claim equivalency. The actual values were ±2.2 dB, ±1.4 dB, and ±2.1 dB for the three earplugs, respectively. Because the 95% two-sided confidence intervals for the prediction errors for each of the earplugs falls wholly within those intervals, this establishes the validity of the F-MIRE measurement system.[23]

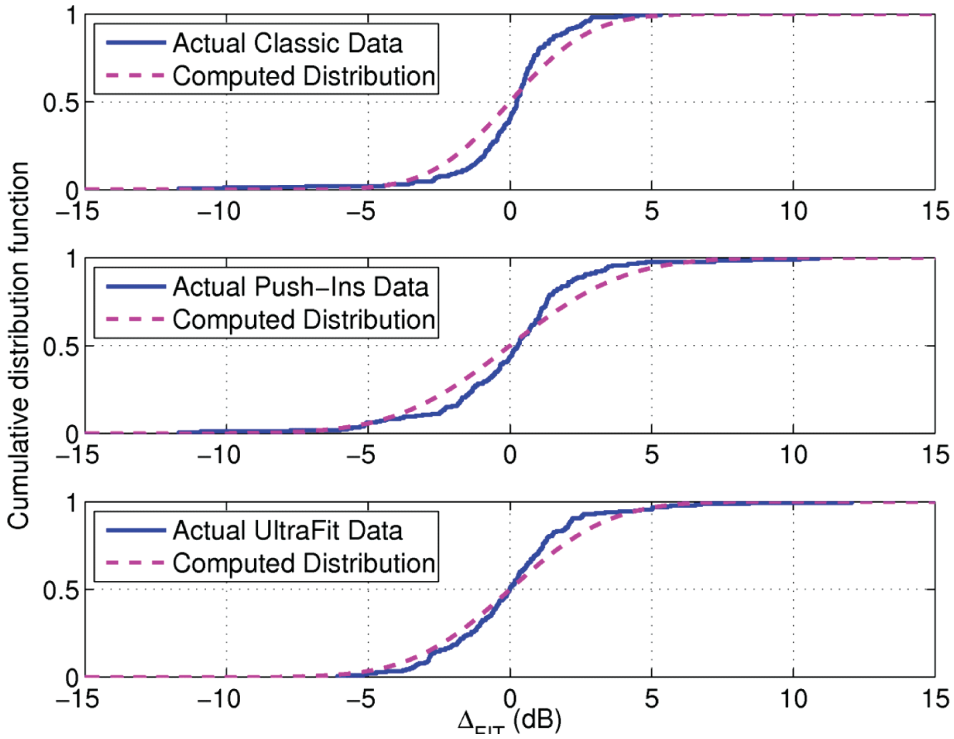


Figure 12: Cumulative distribution curves for the distance-to-mean values for the Classic, UltraFit, and Push-Ins earplugs, from Experiment #2. Solid lines are actual data and dashed lines are computed values presuming a normal distribution

Results II-Fitting variability

Research on F-MIRE testing has revealed that the largest portion of the uncertainty is the fitting variability.[24] In this section, we examine that issue using the data from Experiment #2 in Table 1. For each subject, there are five measurements for each ear. Treating the ears as independent measures provides a total of 40 ears, with five fits (both plug re-fitted to ear and F-MIRE microphone re-fitted to plug), or 200 values in total for each earplug style. We found that the clearest manner of examining the data was in terms of the distance to the mean of the five measurements for each ear, as the best indicator of the true value for that ear is the mean of the five fits on that ear. The distance to the mean is used for each subject, instead of the actual attenuation values they achieved, because, for this analysis, the error on each measurement relative to our best estimate of the true measurement is the value of interest. For each of the fits, the error is computed as the distance to mean (Δ_{FIT}), as described in Equation (5), where n_s indicates subjects index, n_{ear} indicates a subject's left or right ear's index, n_{fit} indicates one of the trials for a given subject, and $\overline{ATT}$ is the mean attenuation across the five fits for that ear, which is measured 40 times (20 subjects × 2 ears). Thus, for each subject, for each ear, and for each measurement, we have:

$$\Delta_{FIT}(n_s, n_{ear}, n_{fit}) = ATT(n_s, n_{ear}, n_{fit}) - \overline{ATT}(n_s, n_{ear}) \quad ...(5)$$

The statistics for Δ_{FIT}, as defined in Equation (5), across the 200 measurements for each product are first examined for outliers in the values collected per ear for each subject by checking for obvious extreme values that would be detected by a technician using the system. The reason to clean the data in this manner is that when the F-MIRE system is actually used, because of the speed of the measurements and ease to conduct a re-test, it is quite likely that users would discard the suspect results and re-test. In fact, this guidance is provided. Thus, when looking for a reliability indicator based on variance in the data, it should be computed for a set of data that presents the type that would be acquired in practice. The only outliers discarded in the current analysis were for the UltraFit earplug. Two values (out of 200) with Δ_{FIT} <-30 dB (corresponding to attenuation values of <2 dB) were removed.

The standard deviations of the culled data sets were then used to assess the standard error of the mean attenuation $\overline{ATT}$. To assess the standard deviation of the values of Δ_{FIT}, given that the five repeated measurements on the same subject cannot be treated as if they were independent, the within-subject standard deviation, defined in Equation (6), is used.

$$s_{\Delta_{FIT}} = \sqrt{\frac{1}{N_s N_{ear}} \sum_{n_s=1, n_{ear}=1}^{N_s, N_{ear}} \sum_{n_{fit}=1}^{N_f} \frac{\left(\Delta_{fit}(n_s, n_{ear}, n_{fit}) - \bar{\Delta}_{fit}(n_s, n_{ear})\right)^2}{N_f - 1}} \quad (6)$$

The standard deviations $s_{\Delta_{FIT}}$ for the culled data sets were 2.2 dB for the Classic, 3.2 dB for the Push-Ins, and 2.8 dB for the UltraFit (two values removed). These within-subject standard deviations are identical to the square root of the within-subject variance that would be computed using a conventional analysis of variance (ANOVA) on the entire 200 values for each earplug.

When the sample mean (the *n* refits of the earplug on one wearer) is used as an estimate of the center of the sampling distribution (the population mean), the standard error of the mean (SEM) can be estimated by the sample estimate of the population standard deviation (sample standard deviation) divided by the square root of the sample size (assuming statistical independence of the values in the sample) using Equation (7).

$$SEM = \frac{s_{\Delta_{FIT}}}{\sqrt{n}} \quad ....(7)$$

Presuming normality, the SEM can be used to establish the 95% confidence interval on estimation of the mean. Accordingly, the data were tested, and were found to fail normality. However, the failure was in a conservative direction, as shown in Figure 12. The values cluster more closely to the mean than in a normal distribution, indicating that when predictions are made of values falling within the bounds of a 95% confidence interval, more values would fall within those bounds than actually predicted. Thus, we proceeded to establish confidence limits in a conventional manner as illustrated in Figure 13.

Figure 13 presents the 95% confidence intervals for the estimation of the true mean value, based on the number of trials (i.e., refits), where the confidence interval of the sampling distribution has been estimated in our study from the within-subject standard deviation of the five measurements on 40 independents ears. Note how the confidence limits quickly diminish on repeat measurements, dropping to approximately 70% of the single-fit error with the addition of just a single re-fit (for a total of two fits), and to 60% of the single-fit error with the addition of two re-fits (for a total of three fits). The error will be a function of the skill of the person re-fitting the plug and of the difficulty of that particular style of plug to be re-fit.

Results III-REAT vs. F-MIRE uncertainity

The results in the preceding section indicate that, on the average, the F-MIRE predictions are reliable indicators of the actual REAT values. However, review of the scatter plots indicates that differences between F-MIRE and REAT for a single measurement on one individual can occasionally exceed 10 dB, and if the scatter plots were presented for individual 1/3-octave bands, such "errors" would be even more frequent. The reason that we put quotation marks around the word "error" is that when a difference is observed between F-MIRE and REAT, one cannot necessarily deduce that the fault lies with the F-MIRE data. There is uncertainty in both the F-MIRE and the REAT measurements and, thus, a difference can also be attributed to an "error" in the thresholds that make up the REAT computation. The scatter

plots are illustrative as the set of data ideally should bracket the one-to-one line; if not, then bias in the F-MIRE estimate is suspected. Figure 11 suggests that there is no bias in the F-MIRE estimates.

To further explore the meaning of F-MIRE vs. REAT differences, we turn to the results of the last of the experiments described in the Procedure section, Experiment #3 on measurement uncertainty that used the Classic earplug. Figure 14 presents the measurement uncertainty for three repeat REAT vs. three repeat F-MIRE measurements on 20 subjects with everything held constant, i.e. one fitting of the plug with the brass nipple inserted for REAT or one fitting of the plug with the probe microphone removed and replaced between F-MIRE measurements or three consecutive open-ear thresholds. The *y*-axis depicts the within-subject standard deviation of the distance to the mean for the open thresholds [here again, the three re-fits on the same subjects cannot be considered to be independent, see Equation (6)], the occluded thresholds, the difference between them (i.e., REAT), and the F-MIRE measures. The F-MIRE values were computed as equivalent binaural attenuation as previously described.

Because we required three sequential occluded thresholds with one given fit of the plug for this part of the experiment (three opens followed by three occluded thresholds), the pairing of open and occluded thresholds was not as clear as is normally the case when open and occluded thresholds are immediately adjacent in time. Thus, for this analysis, we examined the data not only by pairing the first open and occluded values, and the second open and occluded values, and the third, but also on the reasonable assumption that the open- and occluded-threshold variances were uncorrelated, by directly taking the square root of the sum of the squares of those standard deviations. The former procedure is shown by the solid red line and the latter by the dashed red lines. Reassuringly, the answer was equivalent both ways.

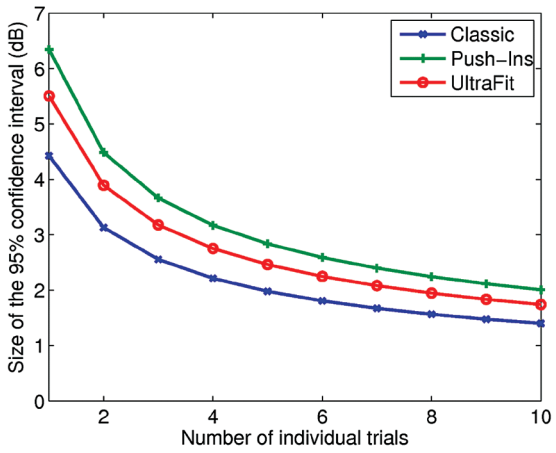


Figure 13: Size of the 95% confidence interval for estimating the true mean value for a given individual as a function of the number of refits (computed as error on distance to overall PAR mean)

With the exception of 8 kHz, where the small wavelengths cause increased uncertainty in F-MIRE measures, the F-MIRE uncertainty is less than or equal to that present in either the open or the occluded subjective thresholds. Moreover, when the difference between the two thresholds is computed, as is required for a REAT measurement, this increases the uncertainty by the square root of the sum of the variances, and the REAT variability becomes substantially larger than for F-MIRE, by more than a factor of two from 125 Hz through 1000 Hz.

Although we have also presented the data in terms of PAR, and similar trends are observed as for the individual octave bands, the frequency-by-frequency data presented in Figure 14 are more compelling. This is because one or two frequencies typically control the PAR computation, as with other single-number ratings such as the NRR, and those depend on the shape of the attenuation curve for a particular product. In this experiment, we used only the roll-down foam earplug, and for that type of product, there is an unusual attenuation characteristic when well inserted [Figure 5], which exhibits as relatively flat attenuation with a dip at 2 kHz due to bone-conduction limitations.[25] The attenuation values at 2 kHz strongly control the PAR and, hence, the Δ_{FIT} at that frequency also strongly controls the variability analysis for PAR. However, for less-deep fittings of the foam plug, or for plugs that exhibit other types of attenuation curves [Figures 6 and 7], the data at other frequencies would dominate and control the standard deviation of Δ_{FIT} for PAR thus leading to somewhat different findings for the relative uncertainties of the REAT and the F-MIRE methods.

Discussion

In developing and validating an F-MIRE approach for

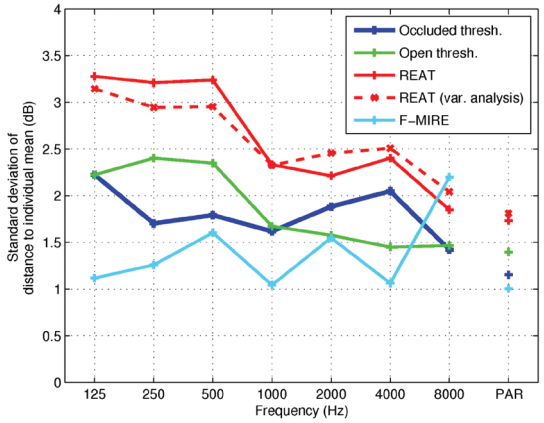


Figure 14: Standard deviation of distance to individual means for 20 subjects tested three times each with a single insertion of a foam plug, for consecutive open and then consecutive occluded thresholds, followed by three insertions and removals of a probe microphone for F-MIRE measurements

measuring hearing protector attenuation in the field, key factors that must be considered include the design of a probe system that is rugged and reliable and that does not substantially affect the use and fit of the plug, the accuracy and precision of the measurements, and guidance for the end user on the meaning and application of the results. The work in this paper has addressed the first two topics; other research and publications have and will examine aspects related to the end-user and application of such systems.[26] In particular, this work explored the acoustical integrity of the probe assembly, the appropriateness and accuracy of compensation factors to convert the F-MIRE noise reduction measurements to equivalent REAT values, the uncertainty of the measurements due to fit variability, and the relative uncertainty between an F-MIRE approach and the acknowledged "gold standard," namely real-ear attenuation at threshold.

The first analysis examined the performance of the probed earplug system. It is important that the probe be small enough and flexible enough not to interfere with the use of the product. However, the probe must also be substantial enough in terms of its inherent transmission loss to assure that the sound levels picked up by the internal microphone element (i.e., sounds from the earcanal traveling through the tube) are not affected by the sound outside the tube. The results presented in Figures 5 through 7 and the accompanying analysis demonstrate that the goal was achieved even for the highest-attenuation product tested, the Classic foam earplug.

Accuracy

The next issue addressed, and perhaps one of greater importance, was the accuracy of the system. Can compensation factors be developed such that the noise reduction measurements derived from the dual-microphone system correctly predict REAT? The data in Figures 8 through 10 explore that issue. We purposely tested the system over a wide range of attenuation values by comparing well-fitted and poorly fitted plugs. The data are shown separately in these figures, and it can be seen that the agreement in octave bands is within about 2.5 dB, with the exception of 2000 and 8000 Hz for the Push-Ins and 500, 2000, and 8000 Hz for the Classic, where the differences are somewhat larger, up to 3.6 dB in the worst case. This agreement is reasonable for the prediction of individual attenuation, especially when one takes into account other variability inherent in specifying earplug performance and the fact that, typically, the descriptor that will be used is an across-frequency rating such as the PAR.

Another avenue to examine the data was shown in Figure 11, where scatterplots were presented in terms of the PAR. The important consideration with respect to that comparison is the realization that disagreements on a given fit-test between REAT and F-MIRE do not necessarily indict the F-MIRE values, as REAT has a substantial inherent variability of its own. Over all of the 80 measurements for each plug it was found that 97–98% of them agreed within $\pm$10 dB, and that the mean error was zero, another indicator of the ability of F-MIRE to predict REAT.

Fit variability

Next, we addressed the fitting variability, as it is clear that fit and attenuation are intimately related. Even with the best intentions, the fit may not be identical each time. Fitting variability was analyzed by using the distance-to-mean metric, or Δ_{FIT} as we called it. This metric was selected as the best indicator of the reliability of a given test, i.e. a comparison of that test with the average of all the tests for that ear. We applied this analysis to our experiment in which 20 subjects fitted and re-fitted both the earplug and the F-MIRE system five times in each ear. The ears were treated as independent measures and hence the 20 subjects generated 40 sets of data, and the results of that analysis are found in Figure 13. For the three earplugs evaluated, the figure shows that if only a single fit is tested, the 95% confidence limits on the prediction of the subject's true PARs, just due to fitting variability, varies from about $\pm$4.4 to $\pm$6.3 dB. It is important to realize that this is not a function of the test system, but rather of the ability to fit and re-fit earplugs. Whether one measures with REAT or any other subjective or objective measurement procedure, this uncertainly is inherent in the fitting of the earplug. The advantage of using an objective F-MIRE system for these types of tests is that the speed of the measurement process, typically 10 s or less, allows testing of multiple re-fits as the principal time cost is then the fitting of the plug and not the time devoted to obtaining the measured results.

Precision

The system uncertainty, or as we call it, the "measurement uncertainty," is described by the results in Figure 14. For this experiment, the repeat measurements were taken with the earplug untouched; we simply conducted a re-test, either another threshold in the case of REAT or another recorded measurement with the microphone re-fitted in the case of F-MIRE. This analysis demonstrates that simply in terms of the uncertainty of the measurement process, F-MIRE is noticeably less variable than REAT by approximately 60% in terms of PAR, and more in terms of the octave-band data from 125 to 1000 Hz. If we compare the standard deviations in Figure 14 to the standard deviations reported in the fitting variability analysis following Equation (5), we note that the values in Figure 14 are 1 dB for F-MIRE measurement uncertainty versus 2.2–3.2 dB for fitting variability, depending on the earplug type. Because the measurement uncertainty is independent of the product tested, it is appropriate to compare the values from Figure 14 with standard deviations for each of the three products in the prior analysis.

Even with the F-MIRE approach, there is uncertainty in the measurements and hence the need to take repeat measurements

to obtain a higher degree of precision. However, as our experiments indicate, the largest portion of the uncertainty is in the fitting variability. Regardless of how the measurements are taken, whether with objective systems such as the one in this study or by subjective field attenuation measurement systems, fitting variability is inherent in the application of hearing protection devices to real ears. This uncertainty can be acknowledged and, if need be, addressed by requiring repeat measurements until the desired precision is achieved. The value of the F-MIRE system is that its speed makes such repeat measurements feasible. Besides, in practice, the repeat measurements serve the added benefit of providing additional practice for employees to improve their fitting technique and enhance the likelihood of obtaining adequate protection from their earplug of choice.

Additional comments

A factor not fully accounted for in the experiments reported herein is validation of the measurement uncertainty with actual employees in a hearing conservation program or with subjects meeting the requirements of the Method B protocol of ANSI S12.6-2008. That was not feasible in these experiments because of the requirements for subject retention and, regardless, the requirements for inexperience would have been abrogated by the multiple re-tests needed in this study. At a future time, we envision testing uncertainty with a revised protocol and Method B-type subjects.

An important issue in designing probes for the foam plugs was to assure that the tubing was sufficiently narrow and soft and that it did not affect the ability to roll down the product for proper insertion. The probes that were developed were found to be quite usable in our experiments. It is unclear however whether the tubing would affect the ability of inexperienced subjects to properly insert the plug. We hope to also test that in a Method B protocol by comparing the REAT values for inexperienced users inserting sealed-tubed plugs and unmodified plugs. However, if the tubing is found to be a problem, it is most likely a "safe" error that would interfere with rather than enhance the ability to insert the plugs. Thus, if the subject can obtain adequate F-MIRE measured protection with the tubed product, she/he will likely do so with the unmodified plug as well.

As the F-MIRE process becomes more widely applied in practice, we anticipate providing guidance on the degree of uncertainty and how it varies with repeat measurements. The recommendations may be application specific and dependent on the level of noise exposure. Additionally, efforts are underway in the ANSI working group S12/WG11 on hearing protectors to develop a standard that specifies how uncertainty shall be measured and reported in field measurement attenuation systems[9] such as the F-MIRE system discussed in this study.

Meanwhile, the hearing conservationist now has available a portable, convenient, quick, and easy-to-use system that can be implemented in programs to improve training and motivation of employees, aid in the assignment of HPDs based on noise exposures, and address other hearing conservation program management and compliance issues.

Address for correspondence:
E. H. Berger,
3M Occupational Health and Environmental Safety Division, E•A•RCAL Laboratory, 7911 Zionsville Road, Indianapolis, IN 46268-1657, USA.
E-mail : elliott.berger@mmm.com

References

1. Royster LH, Royster JD. An overview of hearing conservation practices in the USA. J Acoust Soc Am 1999;105:1009.
2. Gauger D, Berger EH. A new hearing protector rating: The Noise Reduction Statistic use with A-weighting (NRS_A)," a report prepared at the request of the U. S. Environmental Protection Agency, reviewed and approved by ANSI S12/WG11, E-A-R 04-01/HP, Indianapolis, IN, 2004.
3. Berger EH. Assessment of the performance of hearing protectors for hearing conservation purposes. Noise Vib Cont Worldwide 1984;15:75-81.
4. Berger EH. Methods of measuring the attenuation of hearing protection devices. J Acoust Soc Am 1986;79:1655-87.
5. Berger EH. Use of circumaural and supra-aural earphones to measure the real-ear attenuation of earplugs. San Francisco, CA: Am Ind Hyg Conf; 1988. paper 39.
6. Berger EH. Exploring procedures for field testing the fit of earplugs. In: Proceedings, 1989 Industrial Hearing Conservation Conference. Lexington, KY: Off. Eng. Serv., Univ. Kentucky; 1989. p. 7-10.
7. Berger EH. Introducing F-MIRE testing - background and concepts. E•A•R Tech. Rept. 06-29/HP, Indianapolis, IN, Aearo Technologies, 2006.
8. Hager LD, Voix J. Individual field fit testing of hearing protectors – a field-MIRE approach," Conf. American Society of Safety Engineers. Seattle, WA: ASSE; 2006.
9. ANSI. Performance criteria and uncertainty determination for individual hearing protector field attenuation measurement devices. New York, NY: American National Standards Institute; S12.71-201X (draft).
10. ANSI. Methods for the measurement of insertion loss of hearing protection devices in continuous or impulsive noise using microphone-in-real-ear or acoustic test fixture procedures. New York, NY: American National Standards Institute; S12.42-2010.
11. Voix J, Laville F. Expandable earplug with smart custom fitting capabilities. In Proceedings of InterNoise 02, Poughkeepsie, NY: Noise Control Foundation; 2002.
12. Voix J, Laville F. New method and device for customizing *in situ* a hearing protector. Canadian Acoust 2004;32:86-7.
13. Voix J, Laville F. The objective measure of individual earplug field performance. J Acoust Soc Am 2009;125:3722-32.
14. ANSI. Methods of estimating effective A-weighted sound pressure levels when hearing protectors are worn. New York, NY: American National Standards Institute; S12.68-2007.
15. Voix J, Hager LD. Individual fit testing of hearing protection devices. Int J Occup Saf Ergon 2009;15:211-9.
16. Michael PL, Kerlin RL, Bienvenue GR, Prout JH, Shampan JI. A real-ear field method for the measurement of the noise attenuation of insert-type hearing protectors. Rept. No. 76-181, Cincinnati, OH, National Institute for Occupational Safety and Health, U.S. Dept. of HEW, 1976.
17. ANSI. Methods for measuring the real-ear attenuation of hearing protectors. New York, NY: American National Standards Institute; S12.6-2008.

18. Berger EH, Kieper RW, Stergar ME. Policies and procedures manual for the E•A•R / Aearo Technologies E•A•RCAL Acoustical Laboratory re ANSI S3.19-1974 and ANSI S12.6-1997. E•A•R Tech. Rept. 91-41/HP. Indianapolis, IN, Aearo Technologies, 2008.
19. Berger EH, Kieper RW, Peyton DL. Experience with a new ANSI standard for measuring the REAT of hearing protectors (S12.6-1997). J Acoust Soc Am 1999;105:1129.
20. ANSI. Method for the measurement of real-ear protection of hearing protectors and physical attenuation of earmuffs. New York, NY: American National Standards Institute; S3.19-1974.
21. EPA. Noise labeling requirement for hearing protectors. Washington, DC, Environmental Protection Agency, Fed. Regist. 44(190) 40CFR Part 211;56130-147.
22. Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. 3rd ed. New-Jersey: Prentice Hall; 1992. p. 180-1, 199-200.
23. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: The importance of rigorous methods. BMJ 1996;313:36-9.
24. Voix J. Development of a "smart" earplug. Ph.D. Thesis. University of Quebec Montreal, Canada, 2006; Canadiana: 20072127740; ISBN: 9780494208779.
25. Berger EH, Kieper RW, Gauger D. Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways. J Acoust Soc Am 2003;114:1955-67.
26. Berger EH, Voix J, Hager LD. Methods of fit testing hearing protectors, with representative field test data. In 9th Int. Cong. Noise as a Pub. Health Prob. Foxwoods, CT, ICBEN, 2008.

Source of Support: Nil, **Conflict of Interest:** None declared.