A New Hearing Protector Rating: The Noise Reduction Statistic for Use with A Weighting (NRS_A)

E-A-R 04-01/HP

D. Gauger[1]
E. H. Berger[2]

[1]Bose Corporation
phone: 508-766-4206
email: dan_gauger@bose.com

[2]E•A•R / Aearo Company
Chair, ANSI S12/WG11
Hearing Protector Attenuation and Performance
phone: 317-692-3031
email: eberger@compuserve.com

April 29, 2004

Version 2.5

A report prepared at the request of the U. S. Environmental Protection Agency
Reviewed and approved by ANSI S12/WG11
Hearing Protector Attenuation and Performance

ABSTRACT

An obvious and important question to ask in regard to hearing protection devices (HPDs) is how much hearing protection, commonly called attenuation or noise reduction, can they provide. With respect to the law, at least, this question was answered in 1979 when the U. S. Environmental Protection Agency (EPA) promulgated a labeling regulation for hearing protection devices (HPDs) that specified a descriptor called the Noise Reduction Rating (NRR) measured in decibels (dB). In the intervening 25 years many questions and concerns have arisen over this regulation. Currently the EPA is considering publication of a proposed revised rule. This report examines a number of the relevant issues in order to provide recommendations for a new label, new ratings, and a preferred method of obtaining the test results from which the ratings are computed. A wide variety of ratings are reviewed, from the putative gold standard, an octave-band calculation, to simplified ratings employing fewer numbers that can be applied to more common noise measures such as C-weighted or even A-weighted sound levels or exposures. Additionally, the most simplified method of all, namely a class or grading scheme is examined. The conclusion is that a Noise Reduction Statistic for use with A weighting (NRS_A), an A – A' rating computed in a manner that considers both inter-subject and inter-spectrum variation in protection, yields sufficient precision for most situations. Justification for this recommendation stems from consideration of the inter-wearer variation in fitting, the variation in noise spectra, and the accuracy of the basic measurements of hearing protector attenuation and noise-exposure values. Furthermore, it is suggested that to provide additional guidance to the purchaser, two such ratings ought to be specified on the primary package label - the smaller one to indicate the protection that is possible for most users to exceed, and the larger one to indicate the protection that is possible to achieve by individual highly motivated expert users; the range between the two numbers conveys to the user the uncertainty in protection provided. Guidance on how to employ these numbers, and a suggestion for an additional, more precise, graphically oriented rating to be provided on a secondary label (the Noise Reduction Rating, graphical, NRS_G) are also included. Another important consideration is the data from which the new rating is computed. Examination of potential types of data from U. S. or international standards reveals that ANSI S12.6-1997 Method-B data appear to provide the best correlation to field performance and hence the most useful ratings; however, concerns about the reproducibility of Method-B based results led us to also offer an alternative Method-A based value. Since insufficient data are available at this time to clearly distinguish between the two recommendations the need for an interlaboratory study is identified along with suggestions for how it might be conducted.

Contents

0. EXECUTIVE SUMMARY

In the U. S. today there is need for an American National Standard describing a method to compute reliable ratings from hearing protector attenuation data in order to estimate user protection. The lack of such a standard, combined with many concerns that have been expressed with regard to the current Environmental Protection Agency (EPA) hearing protector labeling regulation (EPA, 1979) is the motivation for the work herein.

Many issues are involved in estimating the protection that users achieve while wearing hearing protection devices (HPDs). These include obtaining valid estimates of the HPD's attenuation, as influenced by user training and motivation, as well as the proportion of exposure time during which users actually wear the devices and accurate measurements of the noise exposure in question. Perhaps of greatest concern is the issue of individual variability in the fit and performance wearers achieve. Even with precise computational schemes such as an octave-band analysis of the noise, the issue of variability remains critical. Once predictions are made, one can compute an estimate of the percentage of users in various noises that achieve the targeted protection values, called the protection rate, and use this metric to evaluate the accuracy of various rating systems. For example, if the goal is to protect 84% of the population to a "safe" exposure level, how close does the protection rate approach that desired value?

Numerous rating systems have been proposed in the past 30 years, including publication of two seminal papers. This material was used as the basis for the current research that expanded upon the prior work by introducing new concepts and new data. Ratings of varying complexity were examined, from an octave-band approach, to ratings that must be used with C-weighted sound levels or exposures, from those that work with A-weighted measurements, to those that are simple class or grading schemes. It became apparent that the straightforwardness of what are called A – A' ratings is appealing. Such ratings predict, by simple subtraction from the A-weighted ambient noise levels, the effective A-weighted exposures (A') when an HPD is worn. A – A' ratings, which by their very nature are easier to use and less prone to computational errors, are of sufficient precision for most applications considering the many sources of variability inherent in predicting protection.

An important collateral issue to the development of a rating procedure is the underlying attenuation data from which the rating is to be computed. In the following report we utilized ANSI S12.6-1997 Method-B data for 20 representative HPDs, as well as older style ANSI S3.19-1974 results. We also, albeit with less data available, examined alternatives such as S12.6-1997 Method-A values as well as ones derived from the current ISO 4869-1 standard. Ratings computed from these various laboratory measures of attenuation were compared to ratings computed using data from available real-world studies. We found that Method-B data correlate the best with real-world data, the S3.19 data correlate the worst, and the other two methods fall in between. Although all of the methods could be corrected to correspond, on average across devices, to observed real-world performance, only the Method-B ratings correlate well with the field data and hence similarly rank order the relative performance of different HPDs. Thus it is Method-B data that we emphasized in our analyses, though we also evaluated what corrections might be possible to increase the accuracy of protection rates for ratings based on Method-A data. This enabled us to provide both Method-B and Method-A based labeling recommendations.

Various sets of representative noise data have been published since the 1950s to provide a picture of the occupational noise scene. They originate from around the globe and include both military and specialized environments. Though most of the prior hearing protection analytical studies have based their work on the "NIOSH 100" we included other data sets to make sure that the 50-year old NIOSH data were indeed representative. In large part they were, but in our work we expanded the NIOSH 100 to the Ref300, which include data from a 40-year time span and three countries. We also incorporated specialized Air Force and aviation spectra in order to broaden and validate the findings.

Using the pioneering concepts of Dick Waugh (1976a, 1984) as a basis for our research we analyzed in detail over a dozen different methods of computing ratings of hearing protector performance from the underlying test data. For example, presuming we use a rating such as the current Noise Reduction Rating (NRR) to make predictions for groups of users, and wish to be certain that at least X% of the workers are protected to the criterion we select, how close to X% can we come with the various ratings? Or, what is the average sound level to which under-protected employees will be exposed, given that we would like that value to be minimized? Answering such types of questions allowed us to identify the ratings that best met our goals of simplicity, consistency, and accuracy.

What emerged were two ratings, the Noise Reduction Statistic for use with A-weighting (NRS_A) and the Noise Reduction Statistic, Graphical (NRS_G). Furthermore, in order to provide additional information about the precision of the ratings and to supply better user guidance, we recommend that each of the ratings should be presented as a pair of values representing at the low end what would be possible for most users to exceed in terms of protection, and at the high end, representing what is possible for some motivated expert users to achieve. The same percentiles (80% and 20%) were recommended for the Method-B and Method-A ratings, but in the latter case an additional multiplicative correction factor of 0.71 for the low rating and 0.88 for the high rating were also required to achieve protection rates comparable to those resulting from using Method-B data.

The report summarizes the rationale for the choices we made and provides recommendations on how to implement them, including presentation of the data in a primary label (much like the existing primary label required by law, but incorporating a pair of NRS_A values and new explanatory wording). Secondary and tertiary labels are also suggested that include complete usage instructions, direction in the application of the NRS_G as well as expanded information on the meaning of and use of the range of two values, the avoidance of overprotection, and additional criteria not directly pertaining to noise reduction such as user comfort.

The recommendations conclude with a discussion of the relative merits of the Method-B vs. Method-A based approach, and the fact that insufficient data are available at this time to develop a firm recommendation for one method or the other for rating purposes. Therefore, we call for an interlaboratory study that would produce the results needed to enable development of a firm recommendation.

Annexes are included that provide the actual Method-B, S3.19, Method-A, and ISO data used in the analyses, the noise databases, and sample computational spreadsheet that includes the details of how we compute the ratings contained in this report. The actual hearing protector attenuation data and noise data are available upon request in electronic format.

1. INTRODUCTION

Despite the fact that hearing protection devices (HPDs) have been used for occupational hearing conservation since the early 1950s, widely applied in industry and the military since the early 1970s, and have been subject to a governmental labeling regulation since 1979 (EPA, 1979), there has never been a U. S. voluntary consensus standard developed by the American National Standards Institute (ANSI) specifying how a rating factor should be computed from laboratory derived attenuation data. This is all the more curious since ANSI has published four standards since 1957 (Z24.22-1957, S3.19-1974, S12.6-1984, and S12.6-1997) describing how to conduct real-ear attenuation at threshold (REAT) measurements in a laboratory setting in order to assess the noise reduction that HPDs provide. Because of the absence of consensus guidance, the development of ratings in the U. S. has been based upon government reports and regulations (EPA, 1979; Kroes et al., 1975). This contrasts with the Canadian, European, and Australian/New Zealand experiences in which international or national consensus standards are the defining documents (CSA Z94.2, ISO 4869-2, SA/SNZ 1270).

Another problem exists with respect to the rating and labeling of HPDs in the U. S., namely the fact that a number of reports have suggested that the current Noise Reduction Ratings (NRRs) as mandated by the EPA are an overly optimistic estimator of the actual performance of devices for typical groups of users (Berger et al., 1998; Berger, Franks, and Lindgren, 1996; Berger and Royster, 1996). The difficulties have to do with the REAT data from which the ratings are computed, rather than the details of the computational process itself. However, recent observations suggest that not only should a revised procedure be specified for the development of the attenuation data, but perhaps a more suitable and informative rating can be devised as well. In order to redress this problem, ANSI S12/WG11 (*Hearing Protector Attenuation and Performance*) has turned its attention to the development of a draft standard to describe a new HPD rating, designated herein as the Noise Reduction Statistic for use with A weighting (NRS_A).

To answer the seemingly straightforward question of the sufficiency of the protection that a hearing protector can provide one must specify a method of measuring attenuation over a range of suitable frequencies, include the effects of percentage use time, define the noise exposure of the population or individual in question, and decide upon a computational method for use of those data (i.e., a rating scheme). In each of these areas there is a degree of uncertainty with which we must contend. Though the focus of the current research is the development of a suitable number rating, all four aspects of the prediction problem will be addressed in the discussion that follows.

1.1 Issues in estimating user protection

Valid estimates of HPD attenuation: To begin, we must of course measure the attenuation of the HPD, hopefully deriving data representative for the group of users or specific individuals in question, preferably reflective of their training or skill in use of HPDs. The literature is replete with numerous articles describing the difficulty of accomplishing this task (Berger et al., 1996, Berger et al., 1998), and American National Standards Institute Working Group S12/WG11 spent over a decade crafting a new standard intended to provide improved estimates of field performance (ANSI S12.6-1997). The standard includes two methods, the latter of which (designated Method B) is the one recommended for the most useful estimates of field attenuation, hence, Method-B data are the ones primarily used for the analyses in this paper. Issues regarding the alternative procedure in that standard (Method A experimenter-supervised fit) are also examined in Sections 2.2 and 3.6. Nevertheless, predictions for specific groups or individuals based on laboratory attenuation data are rough estimates at best, unless a fit-check approach is implemented in which the attenuation for the actual users is measured (Berger, 1989; Michael, 1999).

Effects of use time: The next question is the percentage of time the user wears the device when exposed to noise. Since Else (1973) first presented the implications of the energy principal in terms of the diminution of actual protection as the percentage of unprotected wearing time during a workshift increases, many authors have discussed this issue. Clearly it has major impact. In terms of measured protection we often worry about inaccuracies of 2 or 3 dB, yet simply failing to wear a 25-dB HPD for 20 minutes out of an 8-hour shift will reduce the delivered protection by twice that amount (Berger, 2000, Fig. 10.21)[1].

Accurate noise exposure estimates: In order to apply predicted attenuation values we need an estimated noise exposure from which to subtract them. This too has inaccuracies and imprecision. When we account for the accuracy of using acoustical calibrators (± 0.2 dB or greater), the tolerance on microphone frequency response (±1 dB to ±3.5 dB or even more at the frequency extremes), and the sampling size required (for 95% confidence of sampling a worker with an exposure in the top 20% of his or her group, one must sample 8 workers out of 12, or 12 out of 50), it is apparent that even with the best practice one is hard pressed to report measurements with an accuracy of better than ± 2.5 dBA (Earshen, 2000; Royster et al., 2000). In the case of consumer and recreational activities the problems of estimating noise exposures become even more difficult.

It is important to appreciate and place in perspective these sources of error in the overall problem of predicting worker noise exposures as we next we turn to the question of the accuracy of the HPD's noise rating.

1.2 The crux of the problem – individual user variability

In Section 1.1 the issue of developing valid estimates of achieved attenuation was introduced. Typically when this is done, hearing conservationists are dealing with statistical measures and predictions for groups of users. It is illuminating to examine the problem on an individual basis as is done in Figure 1.

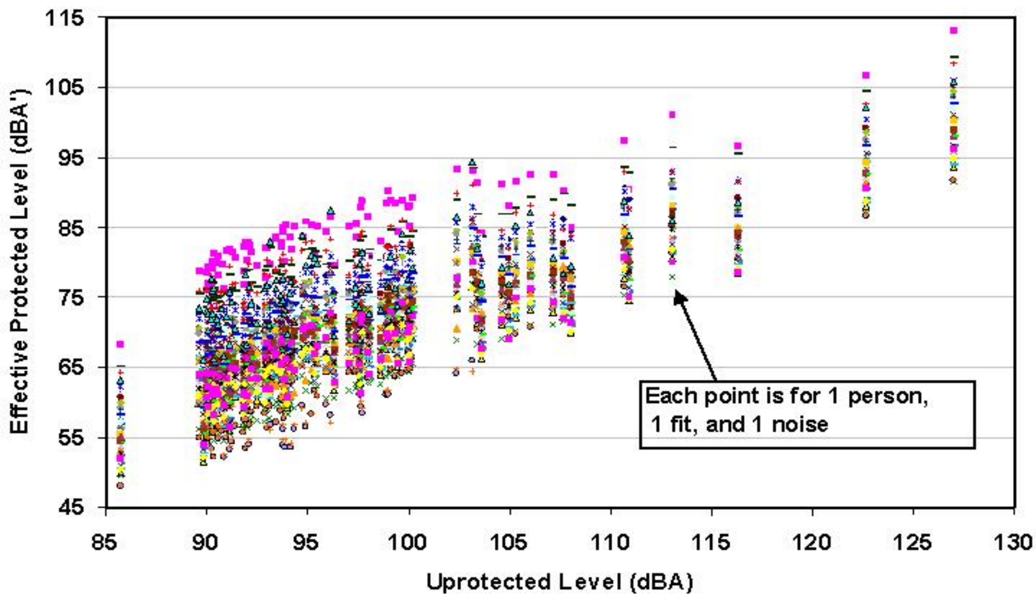
The chart presents the effective protected noise levels for 20 subjects in 100 representative industrial noises. For each subject, an estimate for each of two fittings of the HPD is included. The fitting procedure was according to Method B of ANSI S12.6 to provide an estimate of field expectations. The independent variable (x-axis) is the unprotected sound level in dBA. The dependent variable (y-axis) is the effective protected level in dBA when the hearing protector is worn. For each noise, for each subject and for each fitting of the HPD, the effective protected level was computed using the measured octave-band sound pressure levels of one of the 100 noises together with the octave-band attenuation achieved for that fit by that subject. Each of the 4000 points is the result of one such computation. Since there are 100 noises each with a different sound level, the data group into 100 columns. Within each column are the 40 values computed for the 20 subjects x 2 fits.

The range of effective protected levels for any given spectrum (i.e. any one column in the chart) varies from 18 to 31 dB, averaging 24 dB, across the noises. This indicates that even if the average data from this 20-subject attenuation test were used to make an octave-band prediction for one noise for any one of the subjects, it could be in error by as much as one-half of that range or about 12 dB. Clearly the errors can grow substantially larger when average data from a different group of subjects, or from unrealistic test

---

[1] The estimated 6-dB loss in protection presumes a 5-dB trading relationship between level and duration. Using a 3-dB trading relationship the loss in protection for 20 min. of disuse increases to 12 dB.

Figure 1 – Scatter plot of effective protection achieved by 20 subjects in 100 industrial noises with one hearing protection device.



conditions are used to make predictions. The reason the ranges vary across spectra is because not only do the levels of the spectra vary, but so do their spectral shapes. Since hearing protector attenuation normally varies with frequency so too does the overall noise reduction depend upon the distribution of energy across the noise spectrum.

This wide range of variability must be accounted for in developing a noise rating. In many instances it is this variability that overwhelms all other imprecision in the predictive process.

1.3 Specifying protection and protection percentages

Since noise hazard and permissible noise exposures are nearly always specified in terms of acceptable A-weighted noise levels, or A-weighted time-weighted average (TWA) exposures, it follows that what is generally needed is the A-weighted noise reduction that the hearing protector provides. This is defined as the difference between the A-weighted sound levels (or exposures) when the ear is unprotected, and those same levels (or exposures) effective when the hearing protector is worn.[2] Protected levels are typically denoted by the use of an apostrophe, as in A', which is read as "A prime," and the unprotected values denoted by the same letter symbol without the apostrophe. Symbolically the noise reduction, also called protection, is expressed as,

protection = dBA noise reduction = A – A'

---

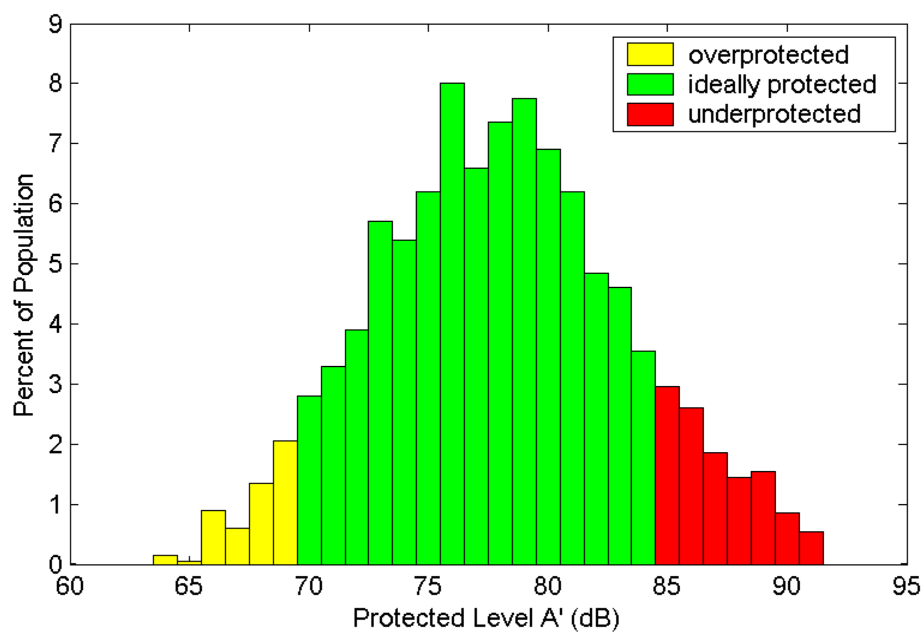
[2] The A' value represents the "effective" sound level when the hearing protector is worn, i.e. the A-weighted sound level at the head center with the listener absent (commonly estimated by the on-the-shoulder measurement with a dosimeter), minus the attenuation of the HPD. This is not the same as the sound level in the earcanal. The earcanal sound level differs from that in the sound field by the transfer function of the open ear. It cannot simply be estimated by subtracting the HPD's attenuation from the A-weighted level (Berger, 1986). However, it is the "effective" values that are required to assess noise hazard, as it is those values that are normally compared to the classical damage-risk curves and permissible exposure limits.

If the HPD provided the same noise reduction to all individuals, and if its attenuation values were equal at all frequencies, then the problem would be quite simple. The HPD would provide the same protection (A - A') for all wearers in all noises. Regrettably such is not the case. Attenuation varies between individuals, sometimes quite dramatically, and since for most hearing protectors, attenuation also varies with frequency, this dictates that even for a given individual, protection will vary from one noise spectrum to another. Somehow, all of this must be captured in a simple rating with a precision appropriate to the underlying data. It makes no sense to construct a complex and "perfectly" precise computational scheme, if indeed the numbers used to generate our computations rest on a bed of sand (see prior section).

The fact that attenuation varies across individuals means that for whatever value of attenuation we specify in a given noise, a certain percentage of wearers will obtain more and others will obtain less than that value. The value we choose to specify will depend on our relative tolerance for underprotection versus overprotection, and our appreciation of the interaction of the variance in the distribution of attenuation values with the precision with which we can make predictions. Figure 2 presents a distribution of A' values for one HPD worn by 20 different subjects in 100 representative noise spectra. Each noise was adjusted using a specified rating (such as the NRR) so that if everything worked precisely as planned, each subject would have experienced an A' of 85 dBA. (More about this procedure appears in Section 2.5.) In fact, as the figure shows, A' varied over a considerable range from below 64 to 91 dBA, but the largest percentage met the target goal of 85 dBA. The variation was due to many factors, only one of which was the imprecision in the rating itself.

Figure 2 illustrates the concept of protection percentage or what Waugh (1984) called the protection rate. It is the percentage of cases in which the desired degree of protection is achieved or exceeded, where a

Figure 2 – Distribution of protection achieved with one HPD worn by 20 subjects in 100 different representative noises.



case is defined as a unique combination of protector, wearer and noise spectrum. In Figure 2 the protection rate, which is represented by the green plus yellow bars, is 88**%.** In this paper we also define the ideal percent protection as the percentage of cases that are not only adequately protected, but also are not protected too much, i.e. not overprotected (see Section 1.7). That group is represented by the green bars (83%) in the figure. It is important to realize that because of the width of most such distributions that the greater the protection rate, the greater will be the percent that are overprotected since the distribution will be moved to the left decreasing the number of red bars while increasing the number and height of the yellow bars (5% in this example as shown in the figure). Depending upon where the median of the distribution falls, one could actually experience a reduction in the ideal percent protection if too severe a goal is selected for the protection percentage. For example in earlier work (Kroes et al., 1975) and in the EPA regulation the goal is a protection percentage of 98%. If this were actually achieved in this example, the overprotected population would increase dramatically to 35% and the ideal protection rate would drop to 64%.

1.4 Types of ratings[3]

Historically, the "gold standard" in estimating the A-weighted noise level at the ear has been an octave-band (OB) computational procedure, analogous to a classical engineering noise control computation for the reduction in OB sound levels from one side to the other of an acoustical barrier (Kroes et al., 1975, ISO 4869-2). Depending upon the lowest frequency of interest, typically 125 Hz, this necessitates computations at seven OB center frequencies. The attenuation values used at each frequency are usually the assumed protection values (APVs), defined as the mean attenuation values from an REAT test such as Method A or B of ANSI S12.6, less a multiple of the standard deviation (SD) across the test subjects. The SD correction presumes that the test data are normally distributed, an assumption that has been shown to not always be true. We address this issue in our analyses.

Typically, the multiple used in the subtraction process alluded to above is one or two times the SD values. Though this computational method seems exceedingly precise, the many assumptions inherent in it give rise to considerable uncertainty in the result. The implicit assumptions in the OB method relate to: accuracy of the HPD attenuation measurements, accuracy of the noise-exposure estimates, use-time in the noise by the wearer being essentially 100%, and normally distributed attenuation values. The first three assumptions were discussed in Section 1.1 and all contain potential and often serious errors. The assumption of normality is typically less problematic, but can influence how we choose to handle the subsequent computations or rating systems. With respect to predicting attenuation for individual users, all bets are off, as there is no way short of personal noise measurements and fit checks (regular individualized measurement of attenuation) to estimate with precision a person's particular noise exposures and attenuation values.

Because of the nearly universal use of A-weighting to specify noise hazard and permissible noise exposures in hearing conservation programs (HCPs), there is a strong desire to be able to rely on A-weighted measurements for hearing protector selection and assignment as well. In spite of the fact that

---

[3] Though it will be explained in greater detail later in the paper, it is important to clarify a point of terminology at this time. We define ratings as numbers computed from the HPD's attenuation values that provide a simplified means of predicting performance; for example, the NRR. The NRR is a type of rating. The method of evaluating the precision and utility of ratings is to create metrics, such as the percentage of the population that achieves the target protection values. Ratings are numbers that are applied to estimate noise exposures, and metrics are numbers that are used to statistically evaluate the ratings.

the existing EPA labeling regulation specifies a rating that should be subtracted from C-weighted noise measurements (see more later), the primary guidance provided in the EPA-mandated wording is to apply the labeled values to A-weighted measurements. Furthermore, one would like to avoid the additional effort involved in acquiring, storing, and utilizing OB noise data solely for the purpose of hearing protector selection. Also consider the fact that the added complexity of using OB or C-weighted measures may confuse the user (Thomas and Casali, 1995) especially in a consumer application. For these reasons, scientists have sought abbreviated ratings that require less data and are easier to apply. These procedures may be grouped into two categories: multiple-number ratings and single-number ratings. For comparison, the OB method may be considered a 7-number (or even 8-number) method corresponding to the seven or eight octave bands that are required for the calculation.

Multiple-number ratings require knowledge of the A- and C-weighted sound levels and the use of one or more noise reduction factors that depend on the particular spectrum shape. Such procedures include Waugh's dBA reduction method (1973), Johnson and Nixon's 2-Number method (1974), the HML method of Lundin (1986) that has been incorporated into an international standard (ISO 4869-2), the current U. S. Air Force 5-number method which is a revision by McKinley of Johnson's method (McKinley, 2001), and a rating computationally similar to the USAF method that we call the NRS_G (Noise Reduction Statistic, graphical), which will be defined later in this paper.

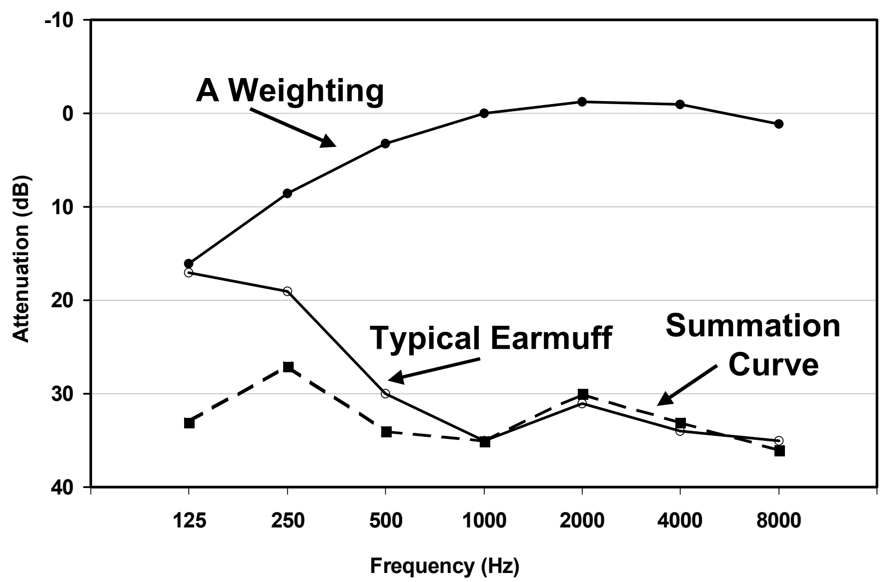
Single-number ratings, as the name implies, require only the measurement of a single sound level, either A- or C-weighted, and the application of a single number or class rating. They have great appeal because of their simplicity of use. Some argue that the single-number ratings that necessitate measurement of C-weighted values are *ipso facto* multiple-number ratings since the A-weighted values will also have to have been measured for purposes of assessing noise hazard or risk. The single-number ratings can be separated into three categories.

- C – A': These ratings tie together C-weighted measurements of exposure and A'. Such ratings include the Noise Reduction Rating (NRR) widely used in North America (EPA, 1979), the Single Number Rating (SNR) embodied in the ISO standard and utilized in Europe (ISO 4869-2), the Sound Level Conversion (SLC_{80}) previously used for rating HPDs in Australia and now used as the basis for the Australian and New Zealand class schemes (Waugh, 1976b; SA/SNZ 1270), Method 2 from the National Institute of Occupational Safety and Health (NIOSH) that is essentially identical to the NRR (Kroes et al., 1975), and the Z factor described in early DIN standards from Germany (Brinkmann and Serra, 1982).
- A – A': These ratings tie together A-weighted measurements of exposure and A'. Such ratings include the use of the NRR with a 7-dB correction as specified by the Occupational Safety and Health Administration (OSHA, 1983) and in the R factor defined in NIOSH Method 3 (Kroes et al., 1975), the NRR_{SF} (indicating subject fit) first described by the National Hearing Conservation Association (NHCA) *Task Force on Hearing Protector Effectiveness* (Berger and Royster, 1996), and the NRF_A, the NRP_A, and the NRS_A defined later in this paper.
- Classes or grades: These ratings assign a class to the HPD based on its attenuation characteristics. They require the use of a look-up table that relates a labeled class or grade to the A-weighted exposures in which the HPD can be used (CSA, Z94.2-02; SA/SNZ 1270). Advantages of classes include their simplicity of use and the fact that fine (and statistically meaningless) distinctions between similar attenuation values are obscured because the classes span a range of attenuation performance. An unavoidable collateral effect is that miniscule differences such as 0.1 dB in the attenuation in a given octave band can cause an HPD to slip over the line from one class to another.

1.5 Basis for C – A' ratings

The reason behind the seemingly odd transformation from exterior C-weighted sound levels to interior A-weighted sound levels that is utilized in the C – A' ratings is not immediately obvious. The origin of this approach lies in the early work of Botsford (1973) who observed, based on OB calculations using data for the typical hearing protector of that era, that the difference between A and A' varied considerably across spectra, whereas the relationship between C and A' was relatively invariant. The explanation is illustrated in Figure 3 for a typical earmuff. The sum of the A-weighting factors and the hearing protector's attenuation values in each OB is approximately constant, within a range of about 5 dB. When this uniform level of noise reduction is applied it behaves like an HPD with the same attenuation at all frequencies, and this is exactly the type of HPD that would provide equal attenuation in all noises regardless of spectrum.

Figure 3 – Demonstration of the C – A' concept.



This line of reasoning led to development of such ratings as the NRR and the SLC_{80}. A limitation is the assumption of the "typical" attenuation curve. The assumption works best for a classical earmuff, less well for properly fitted foam earplugs (popular today), and even less well for the newer types of passive and electronic HPDs that are designed for improved communication capability by delivering relatively flat attenuation across frequency. For such devices the C – A' approach can actually produce less accurate predictions than the conceptually more straightforward and easier to apply A – A' methodology.

1.6 Two seminal papers

Although many authors have suggested methods of developing HPD ratings and others have examined means for determining the appropriateness of such ratings, two seminal papers stand out for their clarity and methodology (Sutton and Robinson, 1981; Waugh, 1984). Their approach forms the basis for our work and the springboard for the additional analyses that we have developed.

Waugh: Waugh pioneered the concept of testing predictions of protection achieved with various HPD ratings by comparing those predictions to "actual" noise reduction values realized in thousands of situations (Waugh, 1976a, then refined in his 1984 paper). As inputs, his procedure requires individual subject attenuation data for one or more HPDs, a mathematical formula defining a rating to be computed from that data, and a set of noise spectra against which to test the rating. For each HPD, the procedure first computes the rating from the attenuation data, then adjusts the level of each noise spectrum so that the result is A' = 85 dBA when the protected level is computed for that spectrum using the rating. Then, the protected level for all combinations of subject attenuation and level-adjusted spectrum is computed using the OB approach, and the large set of resulting A' values examined to determine the protection rate (the percentage of subject/noise combinations below 85 dBA) as well as other statistics. Figure 2 shows a histogram of A' values resulting from such a calculation for one HPD. When using multiple HPDs, the populations of A' values computed for each HPD can be consolidated into one large population representing how the rating performs over a set of devices.

In essence, Waugh's method tests how much a rating can be *trusted* by determining what percentage of subject/noise combinations are actually at a safe level if one uses an HPD in a variety of noise spectra at levels up to the limit that the HPD's rating deems safe. Waugh's method was the inspiration and starting point for the analysis used in the present work. Our extension of Waugh's method is described step-by-step in Section 2.5. In Waugh's 1984 paper, the attenuation data were measured according to the Australian standard in effect at the time, an earlier variant of the current SA/SNZ 1270 subject-fit method that is known to provide useful indicators of field performance. Waugh used data for groups of 15 subjects wearing 30 different protectors (28 earmuffs and 2 earplugs) in 300 different representative industrial noises. This results in a population of 135,000 HPD/subject/noise combinations.

Waugh used three statistics to measure the A' populations he calculated in order to assess the performance of ratings: the mean of the A' values, the standard deviation (SD) of the A' values, and the protection rate in percent. Ideally the protection rate should equal the target inherent in the rating's definition (e.g., 84% if the rating uses an APV equal to the mean less one SD of the attenuation data). The *SD*(A') should be as small as possible indicating that all of the predictions cluster closely about the mean.

In the same paper Waugh repeated his computations with other sets of HPD data (still heavily earmuff biased) and an alternative set of noise spectra, that has come to be called the NIOSH 100 (Kroes et al., 1975). He found similar results. Shortcomings of his paper were the heavy earmuff bias of the data (earplugs are the more commonly used protector in U. S. industry today) and the fact that many of Waugh's conclusions were based on small differences of a few tenths of a decibel in the *SD*(A') values, though no discussion was provided of the practical importance of such small differences.

An essential observation of Waugh was that even the gold standard, i.e. the OB method, has its own limitations in precision. Though this method utilizes OB attenuation values for predictions, those values are based upon data averaged over a *group* of subjects or users; when they are applied to the prediction for any one user, the problem of inter-subject variability rears its head and large variances can arise. Because of the effects of inter-subject variability, Waugh found that the better single-number and multi-number ratings approached the limited precision of the OB method. He concluded that the superior ratings were those that tied C-weighted exposure measurements to predictions of A'. Had Waugh ignored inter-subject variability and simply used a given rating (computed from the average data across subjects) to compare to performance achieved with the OB method (again computed from the average data across subjects) he would have concluded, as is commonly believed, that the OB method provides substantially greater precision than the simplified rating methods.

Sutton and Robinson: Sutton and Robinson conducted analogous computations to those of Waugh using the same 100 NIOSH noises but with a different set of attenuation data (6 earmuffs and 1 earplug). Their conclusions were similar. The principal additional contribution of their work, which we draw upon in our paper, was the concept of the "5th worst overestimation" estimate. Unlike Waugh who based all of his statistics upon the massed data sets and across all noises, Sutton and Robinson found for each protector separately the protection rates for each noise. These values were rank ordered in terms of protection rates and for the 5th worst noise, i.e. the one for which the 5th worst-case of overestimation of protection occurred, the percentage of wearers actually protected in that noise was reported. This analysis assures that the statistics of averages does not obscure problems that may occur for particular HPDs or noise spectra. This concept is embodied in some of the metrics that we describe below.

1.7 Enough protection or too much protection?

Both of the papers described above and much prior work has focused on achieving enough protection for a selected percentage of the workforce, 80%, 84%, 90% or more. However, in recent years attention has also been directed to overprotection and the annoyances and hazards that it creates. In its mildest form it may simply cause HPDs to sound muted or dulled and impair communication with co-workers; in its most extreme form it can create an auditory hazard if important warning and auditory cues are missed because of its effects. Guidance in this realm is difficult to provide in general terms, but one standard does provide a suggested matrix of desirable goals as shown in Table 1 (CEN/TC, 2003). According to this proposal, ideally a protector should provide protection that is from 5 to 10 dB more than is needed to reach the target exposure, but greater than 15 dB of reduction beyond what is needed is deemed overprotection and hence undesirable.

We have embodied this concept in some of the metrics described below.

Table 1 - Assessment of sufficiency of protection according to EN 458.

A' (dB)	Degree of Protection
≥ 85	insufficient
80 to < 85	acceptable
75 to < 80	good
70 to <75	acceptable
< 70	too high (overprotection)

1.8 Differences between this analysis and those of Waugh, and of Sutton and Robinson

Although the fundamental approach employed by Waugh and by Sutton and Robinson is as sound today as it was 20 years ago, there are areas for improvement. The prior papers were written by authors in Australia and Europe where the preponderance of HPDs that are utilized for occupational hearing conservation is more heavily weighted towards earmuffs than earplugs. The data sets they selected reflect that fact. In North America today, the balance is different and it is important that this be reflected in our data. Indeed as our analyses demonstrate, the conclusions differ somewhat depending on the types of products being analyzed.

The type of attenuation data that is utilized also has a dramatic impact on the conclusions. Though Waugh especially, utilized a protocol that provided reasonable estimates of field performance, it was important to select data from today's products tested according to the slightly more refined subject-fit protocol that is now embodied in the current ANSI standard (S12.6-1997) and also reflected in the most current version of SA/SNZ 1270.

Waugh primarily utilized noise data from a set of South Australian measurements developed in the 1960s, though he also included what have come to be called the NIOSH 100; Sutton and Robinson focused their efforts solely on the NIOSH 100. As detailed below we have included those noises but also incorporated a wider range of noises and examined how the particular data set that is selected may affect the results.

Finally, we did not restrict ourselves to ratings that can be easily computed via simple statistical parameters such as the mean and SD of the measurements on a group of test subjects. With the ubiquity of today's PCs it is now easy for any producer or test laboratory, or for most end users to compute substantially more sophisticated ratings based upon distributions of thousands of values, and it was our intent to determine if such ratings might provide beneficial results.

The particulars of our analytical approach are discussed in Section 2.

1.9 Ratings vs. labels

Thus far the discussion has focused on ratings. However, those numbers must be translated to easily digestible guidelines that appear on product packaging as a label. As we discuss later in this document, we find that simple single-number ratings are insufficient to properly describe to users the range of performance likely to be achieved. The question is: to which user are we speaking? Are we trying to provide guidance to a program supervisor who may wish to know in statistical terms what proportion of a population is likely to be protected or is the information intended for an individual who, given a particular fit of the device, wants to know the protection that can possibly be achieved? And once we know if we are targeting groups of users or an individual user, are we targeting typical occupational users with inadequate training and insufficient motivation, or employees in an exemplary hearing conservation program, or consumers who may go to great pains to read instructions and properly select and fit devices, or consumers who blithely buy products and assume they will receive the labeled values of protection without such diligence?

Because of this wide range of possible "consumers" of the label, the approach we suggest in our recommendations is to utilize a dual rating to provide guidance that can be adapted to the needs of the user.

2. DEVELOPMENT OF A NEW APPROACH

With the foregoing in mind we set about developing an analytical approach to the evaluation of HPD rating definitions, and devised metrics with which to assess their performance. The intent was to propose a rating that balances the various considerations described in this section. In the following paragraphs we discuss the data that we utilized and the ratings and metrics that we evaluated.

2.1 The reference set of HPDs

In selecting a group of hearing protection devices for evaluation the following criteria were utilized:

- The devices were drawn from those in the marketplace in the late 1990s and represented the more popular types in use.
- The devices were balanced between inserts/semi-inserts and earmuffs so that there was a sufficient number of each for analysis by type.

- Two “flat-attenuation” devices (one earmuff, one insert) that provide moderate noise reduction were included so that there was the opportunity to assess the newer types of products designed to enhance communication and improve worker acceptance.
- One dual-protection combination was included.
- Within each of the categories devices were selected that represented the extreme as well as the typical range of performance.
- The goal was to represent the range of products currently in the marketplace.

The data used for the analysis were Method-B data (see Section 2.2). The only large body of Method-B data currently available are from Aearo’s E•A•RCAL^{SM} acoustical laboratory, a facility accredited by the National Voluntary Laboratory Accreditation Program (NVLAP) since 1991. Forty-seven (47) tests were available on 22 earplugs, 7 semi-inserts, 15 earmuffs, 2 cap-attached products, and one dual-protection combination. Of the tests about 60% were on E•A•R and Peltor products and 40% were on other brands. The prior efforts of ANSI working group S12/WG11 which wrote the S12.6-1997 standard, indicate good reproducibility between laboratories for Method-B data (Royster et al., 1996), and subsequent informal comparisons between E•A•RCAL and two other laboratories are also suggestive of the fact that E•A•RCAL data are representative. Although there is no guarantee what Method-B data will look like when more manufacturers make them available, it seems reasonable to use the E•A•RCAL results as the basis for the current work.

The selection process, according to the above criteria netted 10 inserts and semi-inserts, 9 earmuffs, and one dual-protection device. The mean attenuation data are summarized graphically in Annex A. In accordance with the test standard by which the devices were evaluated, each earplug was measured on 20 subjects and each earmuff on 10 subjects. Each subject underwent two open and two occluded measurements. The complete individual subject data and summary statistics (means and SDs) are available in a spreadsheet that is available from the authors.

2.2 The type of attenuation data to be utilized

All hearing protectors currently for sale in the U. S. must be tested according to an elderly and withdrawn standard, ANSI S3.19-1974. That standard is known to provide high values of attenuation that are not normally attainable in practice by groups of users. Few today defend those data. The question is what type of test data should be used in its stead. Three options have been proposed. They include either Method A or B from the current ANSI standard S12.6-1997 or the current international standard ISO 4869-1. The merits of the various tests have been debated in the literature as well as at the EPA Workshop held in Washington, DC in March of 2003. Herein we present our reasons for recommending Method B over the alternatives.

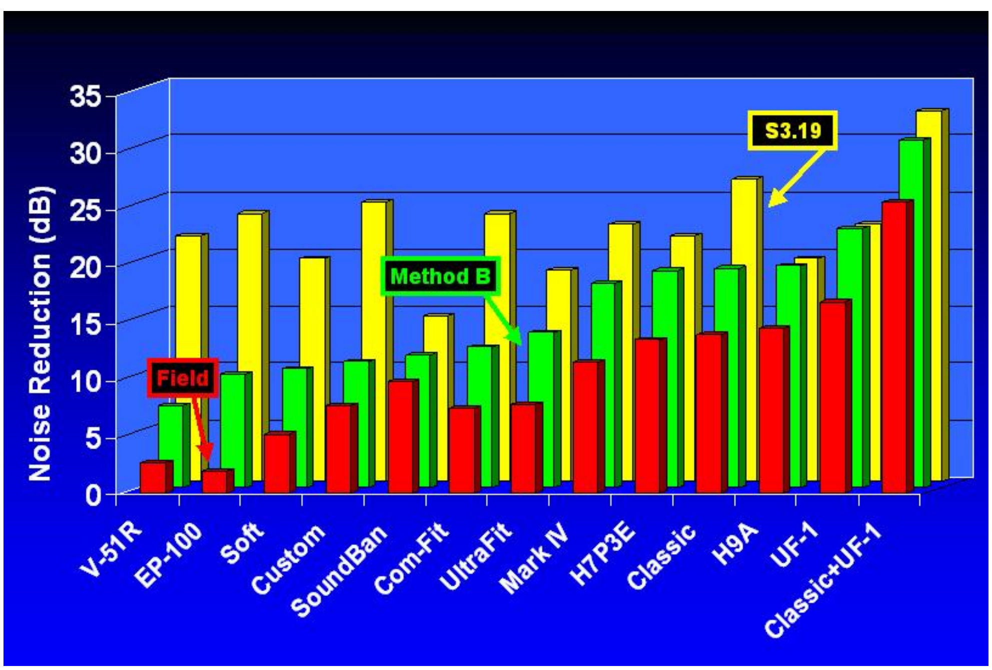
Though the current ANSI standard was promulgated over five years ago, few published Method-B data exist, primarily because the EPA labeling regulation still requires testing to the 1974 standard. However, at least one U. S. laboratory has extensive experience with Method B and it is those data that are used for the analyses herein. Additionally, 120 HPDs have been tested to this method in Brazil by Professor Samir Gerges of the Acoustics and Vibration Laboratory at Santa Catarina, since testing to Method B is required by Brazilian law. Also, the testing conducted in Australia and New Zealand to their current version of SA/SNZ 1270 closely patterns Method B since the Australian/New Zealand working group and S12/WG11 cooperated in the development of their respective standards.

Method B is specifically intended to provide attenuation data that more closely correspond to the performance groups of occupational users obtain in practice (ANSI S12.6-1997, Method B). It also provides values that are representative of the types used by Waugh in his previous studies. Finally, by removing the influence of the experimenter's skill at fitting devices or coaching others to do so, and the effect of the divergent level of skill between resident subject panels in different laboratories, Method B improves interlaboratory reproducibility of attenuation tests (Murphy et al., 2004).

The principal value of Method B is that its results appear to rank order HPDs (from low to high levels of performance) quite similarly to field data (Berger and Kieper, 2000), while also providing a reasonable estimation of the upper bounds of field performance in terms of absolute attenuation (Berger et al., 1998). This is demonstrated in Figure 4 that is adapted from Berger and Kieper. The data shown are for all the HPDs for which there are both field (real-world) and Method-B results available. The S3.19 data are from manufacturers' labeled values, the real-world results are from Berger et al. (1996), and the Method-B data were provided by the E•A•RCAL[SM] laboratory of Aearo Company. Single number ratings are used for the comparison to simplify presentation of the results; use of the octave-band means and SDs would lead to similar conclusions. It should be noted that field attenuation data are susceptible to additional sources of error and variation beyond those described that affect laboratory data, which can lead to underestimates of protection. When multiple field studies were available for an HPD, the figure shows the average of the ratings for the ensemble of studies.

The labeled test data in Figure 4 are from tests using an experimenter fit per S3.19 as required by the current EPA regulation. For a more balanced comparison to the other two sets of data in the figure, which are reported in terms of an A – A' statistic, the NRRs (which are a C – A' type of statistic) have been reduced by 2.5 dB, corresponding to the mean difference in the C – A level in typical industrial

Figure 4 – Comparison of manufacturers' labeled NRRs less 2.5 dB (S3.19), to NRF_A values computed from Method-B data and field (real-world) data. See text for details.



noises (see Section 2.3 and Annex B). We chose this adjustment since if one subtracts the NRR corrected in this way from dBA values (NRR – 2.5 dB), the predictions on the average would be the same as correctly subtracting the NRR (without adjustment) from the C-weighted noise levels. Another difference in the computation of the ratings in Figure 4 is that the NRR is computed with a 2-SD adjustment, whereas the Method B and field data are computed with a 1-SD adjustment that is more apropos for data with realistic SD values. Indeed, if the Method B and real-world data were computed with 2-SD adjustments each of the sets of values would drop similarly and by about 5 dB[4].

The other two sets of data in Figure 4 are the NRF_A values computed for real-world and Method-B data. The NRF_A is an A – A' rating equivalent to the NRR_{SF} plus 1.5 dB as explained in Section 2.4. The figure shows the HPDs ranked by increasing Method-B performance. It is readily apparent that the field data largely follow the same trend as Method B, though on the average the Method-B values exceed the field values by 5 dB. On the other hand the S3.19 data overestimate the real-world results by an average of 12 dB, with a much larger error for plugs than for muffs, and the label test values also rank order devices quite differently than either Method-B or the field data. Looked at in terms of the square of the correlation coefficient, r^2 between Method B and the real world equals 0.95, but drops to only 0.27 when S3.19 is compared to the real world. In other words, Method B laboratory data accounts for 95% of the variance found for different HPDs in the field data, versus only 27% for S3.19. If computed from the individual field study NRF_A values rather than the averages shown in the figure, the r^2 values decrease as expected, with values of only 0.6 for Method B and 0.1 for S3.19.

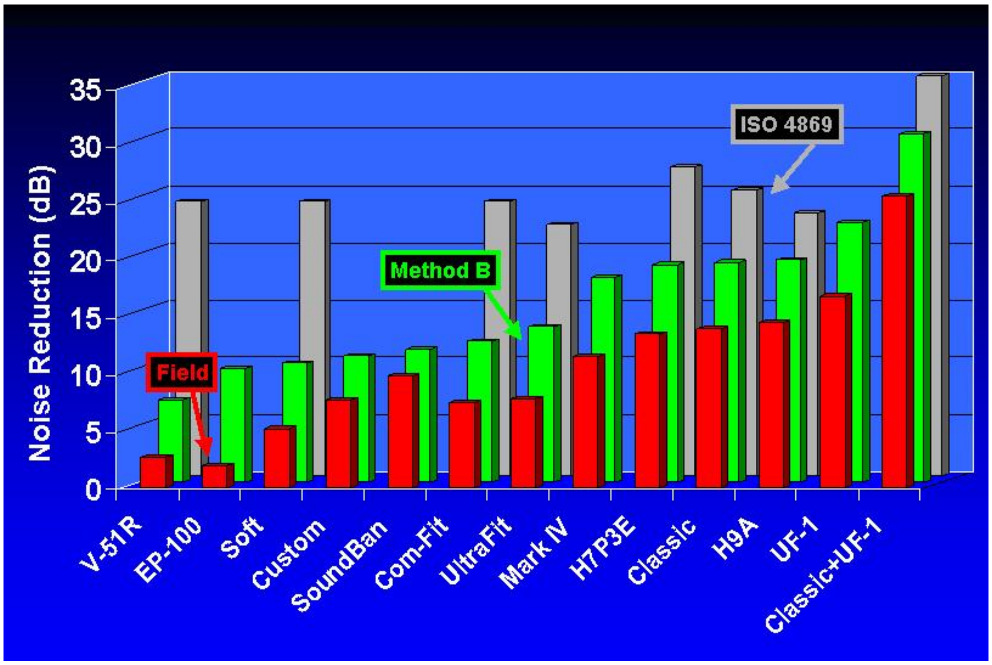
Method A has been proposed as an alternative to Method B. It is defined in the same ANSI standard, is easier for a laboratory to implement, and in the minds of some parties has a greater face validity since it "tests the product and not the subject." Few Method-A data are available at this time, and virtually none are available for those devices on which real-world measurements have been conducted.

ISO 4869 is a third alternative, and in many regards procedurally similar to Method A. There are, however, some subtle and potentially important distinctions such as the requirement in the ISO procedure to instruct the subject to "adjust the HPD for best attenuation consistent with reasonable comfort." Furthermore there is a requirement for the subjects to exercise their jaws and rotate their heads subsequent to fitting and prior to testing. The agreement between ISO 4869 and Method-A data is unknown at this time.

An appealing feature of using ISO data would be the increased uniformity of testing worldwide and the reduction of costs for manufacturers who sell internationally. We were able to obtain ISO data for eight of the field tested devices and created an alternative analysis presented in Figure 5. The ISO data are computed with a 1-SD correction and adjusted by –3 dB so that they are mathematically identical to the NRF_A computations used for the Method-B results. Computed in the same way, the ISO values overestimate the real world by an equivalent absolute amount, as do the NRRs. The correlation to the field data (the r^2 value) for this subset of eight devices is improved to 0.7 for the ISO data versus 0.6 for S3.19 data, whereas the value for Method-B data is 0.997. If one eliminates the particularly high performing dual-protection combination from this calculation, r^2 falls to 0.1 for both ISO 4869 and S3.19 yet remains high at 0.99 for Method B. If one computes r^2 from the individual field study results rather

[4] A direct apples-to-apples comparison of the S3.19 and Method-B attenuation data, computed in terms of an NRR type of single-number (i.e. using a 2-SD adjustment factor) leads to an average difference of 11 dB for earmuffs and 21 dB for earplugs, with the differences for three of the earplugs exceeding 25 dB.

Figure 5 - Comparison of manufacturers' published SNRs less 3.0 dB (ISO 4869), to NRF_A values computed from Method-B data and field (real-world) data. See text for details.



than their averages, one obtains 0.2 for ISO 4869, 0.4 for S3.19 and 0.6 for Method B; if the dual-protection combination is eliminated, the values are 0.1 for ISO 4869, 0.3 for S3.19 and 0.5 for Method B. Clearly, Method B substantially outperforms the other methods in its ability to account for the relative field performance of different HPDs. Accounting for more than 90% of the variance observed, when studies are averaged to obtain a best estimate of the field performance of each HPD, is remarkable.

We are faced with two decisions: first, what attenuation data to use to model real-world HPD-using populations in our implementation of Waugh's method of evaluating rating systems and, second, what attenuation data to recommend as the basis for a new rating. While we have insufficient Method-A data to allow us to judge, we presume that it would correlate to real-world data similarly to ISO 4869 because of the procedural similarities. We also presume that data from additional laboratories and field studies would substantiate the results presented herein. Thus, we choose to answer the first question by using Method-B to model real-world use of HPDs because of its notably better correlation to available field data, as compared to the alternatives shown in Figures 4 and 5 respectively.

Regarding the second question — what attenuation test method upon which to base our rating recommendation — the answer is not as clear-cut. The preferred alternative in our opinion is to again turn to Method B because of its greater ability to account for the field performance of HPDs, in particular and most importantly the relative performance they demonstrate. We disagree with those who say that Method B rates the subject and not the product, saying rather that Method B measures the product, as influenced by its ease of use and quality of the instructions provided. Just as Method-B data can be affected by the varying quality of subjects, some of whom diligently read and attempt to follow instructions, real-world users of HPDs vary in their diligence at putting into daily practice training they may (or may not) have received in the use of HPDs. However examination of the increasing amounts of Method-B data that are now available suggests that although its variability both within and between laboratories is similar to more controlled ISO and Method-A procedures in terms of the absolute deviations in decibels, that its variation is greater as a percentage of the rated values, since Method-B rated values are by their very nature smaller in magnitude (see Figures 4 and 5). This could unduly sway purchasing decisions and cause hardship for manufacturers and product designers. Thus we endeavored to develop the best possible ratings for both Method-B and the more-controlled Method-A test data. We also recommended that supplementary Method-A and Method-B data from additional laboratories are needed before the second question can be answered with confidence. In Section 4.6 we describe an interlaboratory study that should be conducted to provide a stronger foundation for this important decision.

In consideration of the arguments concerning the suitability of Method B, we offer an analogy to the EPA's fuel economy rating. An automobile's gas mileage can be measured under optimal conditions, at a steady speed chosen for maximum efficiency. Arguably, that is the performance of the product — the car in isolation from other factors. The expectation of consumers is that the fuel economy rating reflects how they use their cars, in varying driving conditions, speeds, terrains, and amounts of acceleration. The EPA attempts to simulate the latter in the laboratory with its present method of measuring fuel economy by having cars driven on dynamometers over ostensibly representative speed and acceleration profiles. Some consumers may do better by being diligent about using moderate speed and acceleration. Most will do worse because of their inattention to these details plus the vagaries of varying traffic conditions, road surface quality, wind, etc. The laboratory fuel economy test attempts to pattern real-world operation of vehicles, just as Method-B has been shown to do with HPDs.

If adopted, a Method-B based rating would reflect not only the attenuation performance of the device, but its ease of use as well. To the extent that the market demands it, we would hope that such a measurement and rating paradigm would motivate manufacturers to offer easier-to-use and hence more consistently performing and predictable HPDs to the benefit of industrial, consumer and military users.

2.3 The noise data

The noises that were utilized for our analyses were taken from classic reports available in the literature, combined with one set of data provided by the first author.

- NIOSH 100 – These noises were selected by NIOSH (Kroes et al., 1975) in the 1970's from the work of Karplus and Bonvallet (1953) who measured 579 noise spectra from various industries. NIOSH selected the noises based on standard industrial codes (SICs) to be representative of general industry. Subsequently, the noises have been extensively utilized for many types of predictions and have become the most widely used reference industrial noise database, worldwide. However, today the data are over 50 years old.
- South Australian (SA) 300 – These noises were selected by Dick Waugh for his hearing protector analyses, by arbitrarily selecting 300 noises from McQueen et al. (1969) who reported 615 noises. It is unknown what, if any, statistical approach was used in the initial sampling of the noises to assure their representativeness.
- New Zealand (NZ) 230 – These noises were collected from 1987 – 1994 in an attempt to be representative of New Zealand industry (Backshall, 2000). Of the 282 measured noises, only the 230 that provided levels of 85 dBA or greater were used in this analysis.
- Air Force 50 – This noise database was published by Johnson and Nixon (1974) and has become the standard used by the military to represent spectra of jet engines, helicopters, and ground

auxiliary equipment, as well as many industrial type situations found in the U. S. Air Force. The noises were selected from about 700 noises that were available at the time.
- Civil Aviation 20 – These noises were collected by the first author in the 1990s to be representative of civilian aviation aircraft in common use, and with the specific intention to capture the widest range of spectral variation that could be experienced, especially with respect to noises with substantial low-frequency content. All measurements were taken in either the cockpit or cabin of aircraft in flight.
- Reference 300 – the development and rationale for this noise database, consisting of 100 noises each from the NIOSH 100, SA 300, and NZ 230 databases is described in Annex B. It is intended to be a reference set of industrial noises that spans both geography and time. It is that noise database that was principally used for the analyses that follow.

The noises are examined in terms of the distribution of their spectra and the development of the Reference 300 (designated Ref300) in Annex B.

2.4 The ratings

The ratings we have chosen to include in our analysis encompass the range of rating types described in Section 1.4 and include the major ratings in use today. To this are added three new ratings we have developed and tested, one of which we recommend in Section 4 as the basis for a revised primary label, and another of which we recommend as an improvement to the secondary label.

Goals: The rating we recommend for the primary label, NRS_A, and its secondary label companion, NRS_G, are proposed in an effort to meet the following goals that we suggest should characterize an ideal rating.

a) The rating for the primary label should be of the A – A' type so that it can be subtracted from A-weighted noise levels or TWAs to estimate the protected level since, most commonly, only A-weighted data are available to describe a noise environment. In contrast, the NRR is a C – A' rating that requires the user to remember to subtract a correction in the absence of C-weighted noise data, a source of error in application of the rating (Thomas and Casali, 1995). Using an A – A' rating is most important to consumers who very likely have no knowledge of A or C weighting. As an adjunct to the primary label, the secondary label should present the A – A' protection for different C – A values of noise in a simple, easy-to-use fashion to increase accuracy when C-weighted data are available. The consistency of protection rate for varying noise spectra for different rating types (A – A', C – A', multi-number) is examined in Section 3.4.

b) The rating should be computed from ANSI S12.6-1997 Method-B attenuation data for the reasons discussed earlier in Section 2.2. Section 3.3 compares the protection performance of ratings based on Method-B to the current NRR. We also present an alternative of basing the rating on Method-A or ISO 4869-1 data; Section 3.6 compares the protection performance of Method-B ratings to this alternative.

c) The "single number" rating for the primary label should actually consist of two numbers (low and high values) that convey to the user the range of performance that an HPD may be expected to provide. A single number creates a false impression of precision and encourages an unwarranted focus in device selection on slight differences in rating values. In Section 4 we describe how the two-numbered rating we recommend should be used to determine which HPDs are sufficiently protective for a given noise environment. The new ratings that we introduce in the following section, the recommended NRS_A and NRS_G, and one we have rejected (NRP_A), provide both low and high values.

d) The rating should perform consistently for different HPD-types, ideally coming close to the targeted protection rate for any HPD and noise to which it is applied. Similarly, the range between the two numbers for the primary label rating should depend on both the inter-subject variation in attenuation as well as the variation in protection with noise spectrum. Traditionally single-number ratings have addressed the influence of spectrum on attenuation either by use of a constant adjustment (for A – A' ratings) or by encouraging the user to apply a C – A' rating when dealing with atypical spectra. Both of these approaches presume a traditional, earmuff-like sloped attenuation response. Today, "flat" or moderate attenuation devices exist for which these assumptions no longer hold. A rating that is not built upon assumptions about the shape of the attenuation response and that conveys the range of protection an HPD offers across both typical user fit and noise spectra should encourage innovation and development of HPDs that offer more predictable, natural sounding protection. The rating we recommend for the primary label accomplishes this for typical industrial spectra, the companion rating for the secondary label (NRS_G) provides the greater information needed to determine protection in noises with a C – A values that exceed those typically found in industry. This issue is explored in Section 3.4.

e) <u>The rating should target a protection rate of 80 or 90%</u> because these are easy for users to relate to ("4 out of 5" or "9 out of 10" people will exceed the lower value). The NRR targets 98% protection by subtracting two SD from the mean attenuation; in using subject-fit rather than experimenter-fit data a one-SD adjustment corresponding to 84% protection has more typically been used (e.g., the NRR_{SF}). The SNR example given in ISO 4869-2 corresponds to an 80% protection target by subtracting 0.84 SD from the mean. The tables in Section 3 present results at both the 80 and 90% targeted protection rates for the low value, and also for one high-value candidate, 20%. These are achieved either by means of normal statistics or by direct computation of percentiles on large sets of protection values, as discussed in Section 2.6. The pros and cons of basing the rating on the various protection rates presented are discussed in Section 4.

f) Attenuation data, particularly data obtained using Method-B, sometimes exhibit non-normal behavior, sometimes with values concentrated near two "modes." Accordingly, <u>the rating should balance minimal use of the assumptions of normal statistics with minimal sensitivity of rating values to changes in attenuation data</u> that can be expected with repeated tests within one laboratory or between different laboratories. Method-B has been shown (Royster et al., 1996; Murphy et al., 2004) to achieve statistically better inter-laboratory reproducibility of attenuation data than S3.19. This is in part the result of higher inter-subject variation in the results. Since the mathematics defining a rating can cause the resulting values to depend to varying degrees on changes in the attenuation data, it is important that the rating value be tolerant of inter-subject variation. The results in Section 3.2 explore this issue.

Ratings: The ratings for which we report results are listed below. Each is designated by an abbreviation and briefly described. With the exception of the NRR and its variants (which require S3.19 data), all ratings may be calculated using Method-B, Method-A or ISO 4869-1 REAT attenuation data over the octaves from 125 Hz to 8 kHz. In the list, *Q* refers to the targeted protection rate, the percentage of the population exceeding the rated noise reduction; *Q* = 84% (i.e., minus one standard deviation) has been the value most commonly reported when using Method-B and ISO 4869 data. In the remainder of the paper we explicitly identify the *Q* value for a rating, generally via a subscript. The values of all ratings for the twenty reference HPDs are given in Section 3.1 and the results of our implementation of Waugh's method of assessing rating performance are given in the remainder of Section 3.

Single number, C – A' type, apply by subtracting the rating from the noise's C-weighted value:

NRR ....Present EPA Noise Reduction Rating computed from S3.19 data. Defined as C – A' – 3 in pink noise (noise with equal energy in each octave) computed with an APV corresponding to Q = 98% assuming normal statistics.

SNR_Q............ISO 4869-2 rating defined as C – A' in pink noise.

Single number, A – A' type, apply by subtracting the rating from the noise's A-weighted value:

NRR–7 ....Defined as NRR minus 7 dB, the present government recommendation for how to adjust the NRR for use with A-weighted noise measurements (Kroes et al., 1975).

$NRR–7_{OSHA}$....OSHA recommended derating of the NRR, defined as (NRR – 7) / 2 (OSHA, 1990).

$NRR–7_{NIOSH}$...NIOSH recommended derating of the NRR, defined as NRR*D – 7 where D = 0.75 for earmuffs, 0.5 for foam earplugs and 0.3 for other earplugs (NIOSH, 1998).

NRR_A.............NRR minus 2.5 dB, the mean C – A for industrial noise; the NRR adjusted for subtraction from A-weighted noise data as used in Figure 4 (Section 2.2).

$NRF_{A,Q}$...........Noise Reduction Factor for A-weighted noise. Defined as C – A' – 3 in pink noise using an APV_Q determined using normal statistics. $NRF_{A,84}$ – 1.5 = NRR_{SF} = SNR_{84} – 5 (Berger and Royster, 1996); we adjusted $NRF_{A,84}$ to be 1.5 dB higher than the NRR_{SF} to bring the calculated protection rate as close as possible to the target of 84%.

Class ....A modification to the CSA Z94.2 HPD grade system. The rated noise reduction (NR) is defined as follows: NR = 20 for $SNR_{84} \geq 20$, NR = 15 for $SNR_{84} \geq 16$, NR = 10 for $SNR_{84} \geq 12$ and NR = 5 for $SNR_{84} \geq 8$. The SNR thresholds have been adjusted downward 2 dB relative to the values in Z94.2 to come as close as possible to an 84% protection rate.

$NRS_{A,Q}$ ....Noise Reduction Statistic for A-weighted noise. Defined by computing A – A' for all combinations of subject-mean[5] attenuation and each of the NIOSH 100 noise spectra, and then computing the mean and SD for the ensemble of values. The low and high protection ratings are calculated from the resulting values using normal statistics and two targeted protection rates. In this report we examine two possibilities for the low value (Q = 80 or 90%) and one for the high value (Q = 20%).

$NRP_{A,Q}$ ....Noise Reduction Percentile for A-weighted noise. Defined by computing A – A' as in the case of NRS_A then finding the low and high protection ratings from the resulting values by direct calculation of the percentiles of the ensemble of values.

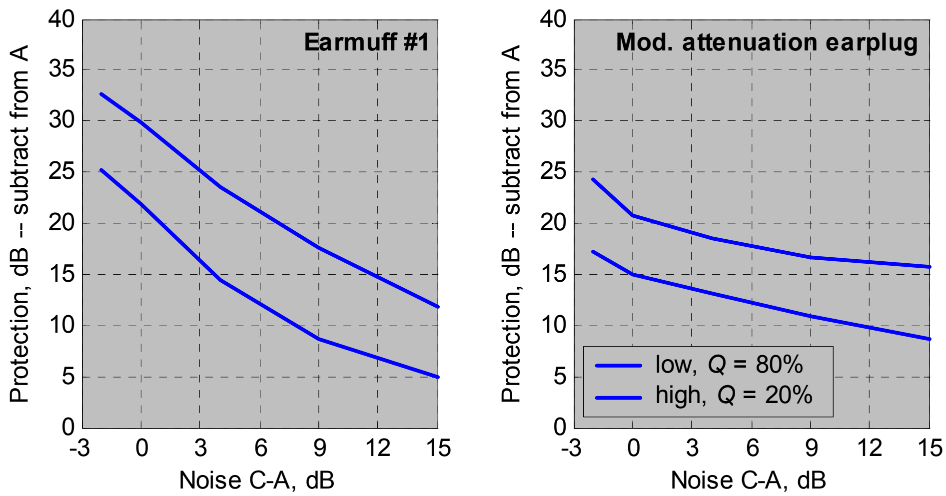
Multi-numbered ratings:

OBN_Q ....Octave-band method (the "gold standard") using normal statistics. The APV at each frequency is the mean of the attenuation data minus the appropriate number of standard deviations to achieve the target protection rate Q. A' is calculated from the resulting APVs and the octave-band noise spectrum in which the HPD is to be used.

$NRS_{G,Q}$ ....Noise Reduction Statistic, Graphical. Consists of low and high protection ratings at five different noise C – A values (-2, 0, 4, 9, and 15 dB). These values are determined by computing the protection for each subject in each noise in a set of widely varying spectra, finding the protection values for each noise corresponding to the low and high protection rate targets, then fitting four line segments to the resulting protection values. Applied by reading the protection from the graph corresponding to the C – A value of the noise (Figure 6) or, for each protection rate, interpolating between the adjacent low and high protection values for the two adjacent C – A values in a table.

[5] Subject-mean attenuation is the mean across trials for each subject in the REAT data set.

Figure 6 – Example of NRS_G Graphical Presentation



Annex C provides more detailed information on the calculations used in this report, including a table that explicitly defines each rating with a mathematical formula. The calculations are illustrated for NRS_A and NRS_G using a Microsoft® Excel® spreadsheet that is available electronically from the authors. All values computed in Section 3 were actually calculated using a set of functions written in Matlab® technical computing language (The MathWorks, 2000). The Matlab code used is available from the first author; Annex C also contains a description of what this computational environment includes.

Note that NRS_A relies on normal statistics like existing rating methods, but applies them to a different set of data. Existing ratings apply normal statistics to the attenuation data for a group of subjects to determine an APV. This APV is subtracted from a defined spectrum (pink noise), and the single-number rating is computed from the resulting protected octave-band levels. This fails to factor into the result the influence that the shape of the HPD's attenuation response can have in different noise spectra and the degree of correlation in the attenuation between different frequency bands. In contrast, NRS_A computes the protection for all the subjects tested in the 100 NIOSH noises, resulting in a set of 1000 A – A' values for earmuffs (100 noises times 10 subjects for Method-B) and 2000 values for earplugs (20 subjects).

Existing single number ratings also assume that the influence of noise spectrum is captured in either the use of a C – A' form of the computation, as illustrated earlier in Figure 3, or by means of a constant adjustment in the case of A – A' ratings like the NRR_{SF}. The NRS_A makes no such assumptions about the shape of the attenuation but simply computes the protection in the NIOSH 100 noises representative of the industrial workplace. By reporting two values for different target protection rates *Q*, this approach causes the range between the values to be a function of the uncertainty in protection due to both inter-subject variation in attenuation and attenuation variation with frequency, in accord with goal (d) stated earlier. Finally, note that NRS_A calculation is based on the subject-mean attenuation values consistent with the calculation of standard deviation prescribed by ANSI S12.6-1997. We also use this subject-mean attenuation when computing the SNR from Method-A and Method-B.

The NRS_G presents the protection as a function of noise C – A value like the HML method defined in ISO 4869-2 (Lundin, 1986), and a five-number method used by the U. S. Air Force (USAF) (McKinley, 2001). The key improvement we offer is that the results should be presented graphically on the secondary label. Properly using the HML requires arithmetic that, while straightforward, is intimidating for many people. Our belief is that a graph such as the one shown in Figure 6 that utilizes one parameter such as C – A to characterize the noise will make using a multi-number rating easier for the majority of people. The HML or USAF method could of course be presented graphically as well; however, we chose for reasons of consistency to instead define a multi-number rating whose calculation parallels that for our primary label recommendation NRS_A by applying individual subject attenuation values, rather than an APV, to each noise. Also, compared to the USAF method, the NRS_G is based on a larger set of noises (the NIOSH 100, Air Force 50 and Civil Aviation 20) and a different set of C – A values that define the ends of each line segment on the graph.

The list above and the results in Section 3 do not show every rating we examined. For example, C – A' variants of the new NRS_A and NRP_A ratings are obvious extensions: we report no results for them because we deemed having an A – A' rating to be a paramount consideration and our results showed no overriding advantage to a C – A' rating such as the SNR. In earlier research, WG11 examined an octave-band rating based on percentile statistics (OBP) and a direct percentile computation analog to NRS_G we called NRP_G. We rejected OBP for reasons discussed in Section 2.6 and NRP_G because of disadvantages to direct percentile computation reported in Section 3.2.

2.5 The Waugh analysis and metrics

The analysis on which we base our rating recommendations is an extension of the work by Waugh described in Section 1.6. In honor of the originator of this approach, we refer to this as a "Waugh analysis" and the set of A' values generated as the "Waugh population." We refer to statistics computed on the Waugh population as metrics; these measure the performance of a hearing protector rating in different ways.

We outline below the steps of the Waugh analysis as we have implemented it.

1) Assemble a set of HPDs and their associated attenuation data to be used to test a rating. Attenuation data are needed to calculate the rating as well as to represent the performance that a group of individuals obtains when using the HPD. In Section 3 we use S3.19, Method-A, Method-B, and ISO 4869-1 REAT data for computing the rating, and either Method-B or field data for computing the Waugh population. Method-B data[6] are used for computing the Waugh population in most of our results because, as shown in Figure 4, they correlate better with field attenuation than the other methods.
2) Assemble a set of octave-band noise spectra to represent environments in which to compute protection. In most of our results we use the Ref300 noises for this purpose because it is a large set of spectra representative of industrial noise environments in the US. We also use the Ref300 rather than the NIOSH100 because some of the ratings to be tested use the NIOSH100 in their definition, so testing them with the same set of noise spectra seems biased in their favor. In some comparisons we also use a collection of 70 aviation noise spectra (the AF50 and CA20, Annex B) in order to test rating performance in environments quite different from industrial noise but where, nonetheless, users would assume a single number rating was meaningful.

[6] We use subject-mean attenuation (mean across trials) for computing the Waugh population because it is the best estimate of what each subject would achieve on average over an extended period of time.

3) For each HPD in the set
   a) compute the value of the rating[7] to be tested, and then,
   b) for each noise in the noise database, estimate the A' value using the selected rating and
      - Shift the spectrum up or down the same amount in each octave so that the A' value estimated by the rating is 85 dBA.
      - Compute the actual A' value using the OB method for each subject in the level-shifted spectrum using the individual subject-mean attenuation values.
   c) Repeat step (b) for each noise and assemble the A' values computed into the Waugh population for each HPD.
4) Repeat step (3) for each HPD, assembling the A' values into the complete Waugh population for the set of HPDs.
5) Compute the metrics listed below from the Waugh population.

The procedure outlined above yields a large number of A' values upon which to compute the metrics. For example, when doing the Waugh analysis with all 20 HPDs from the database described in Annex A the procedure uses nine earmuffs with ten subjects each in the attenuation data and ten plugs plus one double-protection combination, each with twenty subjects. The resulting set of Waugh attenuation data represents a total of 310 sets of subject attenuation responses. The procedure combines each of these with each noise from the Ref300 database, yielding a total of 93,000 protected levels[8] from which to compute the metrics for each rating.

As mentioned in Section 1.6, the essence of the Waugh analysis is to analyze a large set of individual protected levels in different noise spectra that result from trusting a rating's simplified representation of an HPD's attenuation in order to assess how much that trust is warranted. The metrics we compute from the Waugh population to guide us in that assessment are listed below**[9]**.

$P_{protected}$ .... The protection rate — the percentage of the Waugh population for which A' is less than 85 dBA, the level to which each noise is shifted based on the protection estimate provided by the rating. Ideally this should equal the target protection rate, *Q*, for the rating.

P_{ideal} .... The percentage of the Waugh population that is neither underprotected nor overprotected in accordance with Table 1 ($70 \leq A' < 85$ dBA).

A'_{under} .... The underprotected equivalent level; i.e., the A-weighted noise level equivalent to the average noise dose for the members of the Waugh population that are underprotected, calculated using a 3-dB exchange rate.

SD(A').......... The standard deviation of the protected level A' across all members of the Waugh population. This measures the breadth of the distribution.

---

[7] When presented on labels, rating values are almost always rounded to the nearest whole number value and some standards explicitly state this is to be done. We did not do this because we wanted the underlying mathematics of the rating to determine our results. For some of our metrics, rounding would have significantly affected the answers though not the conclusions reached.

[8] The 20-HPD database is nearly evenly balanced between earmuffs and earplugs. The greater number of subjects used for earplug tests yielded 71% of the noise exposures to be earplug users. This corresponds more closely with the balance of earplug to earmuff use in U. S. industry.

[9] Two of these metrics, SD_S(Rating) and $SD_N(P_{prot.})$, were suggested by William Murphy of NIOSH as ways to assess the precision and accuracy of rating methods, respectively (Murphy, 2003).

$SD_N(P_{prot.})$ .... The standard deviation of the value of $P_{protected}$ computed across noises, after combining all HPDs and subjects for each noise in the set of spectra used. The smaller the value the more consistently a rating describes an HPD's noise reduction in different noise spectra.
$WC_N(P_{prot.})$ ... The worst-case (lowest) value of $P_{protected}$ observed across all HPDs for any noise.
$WC_N(A'_{under})$. The worst-case (highest) value of A'_{under} observed for a noise with the worst-case value of $P_{protected}$.
$SD_H(P_{prot.})$ .... The standard deviation of the value of $P_{protected}$ computed across HPDs, after combining all noises and subjects for each HPD. The smaller the value the more consistently and fairly a rating performs for different types of hearing protectors.
$WC_H(P_{prot.})$ ... The worst-case (lowest) value of $P_{protected}$ observed for any HPD in the set being analyzed across all noises.
$WC_H(A'_{under})$. The worst-case (highest) value of A'_{under} observed for a hearing protector with the worst-case value of $P_{protected}$.

One more metric we use is not computed from the Waugh population but rather solely from the attenuation data provided for the set of HPDs.

SD_S(Rating) .. The standard deviation of the rating value across sets of subjects. This is computed using the "bootstrap" method described in Section 3.2 (Murphy, 2003) rather than from the Waugh population. It measures the uncertainty in the rating value as a result of the attenuation distribution across subjects. A smaller SD_S(Rating) is desirable because it indicates a greater likelihood of reproducibility of the rating value during repeated testing within one laboratory or in different laboratories.

2.6 Normal statistics vs. percentiles

Existing rating methods rely on normal statistics (means and standard deviations) to model the variation in the attenuation data to determine an APV that should result in achieving the targeted protection value *Q*. Table 2 shows the number of standard deviations by which to adjust the mean to achieve various protection targets.

Table 2 – Percentile values for normally distributed data.

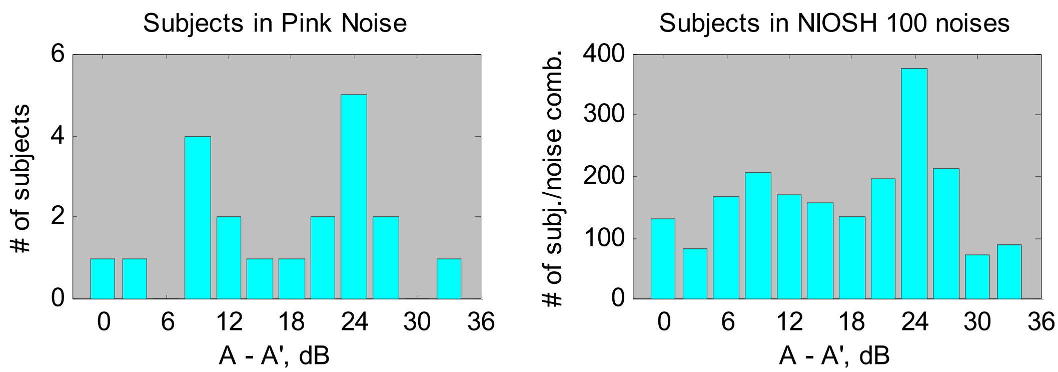
Target Protection Rate *Q*	98%	95%	90%	84%	80%	50%	20%	16%	10%
Standard Deviations	-2.00	-1.64	-1.28	-1.00	-0.84	0.00	+0.84	+1.00	+1.28

A topic of discussion in WG11 for several years has been the occasional non-normal character of attenuation data, in particular for earplugs. Non-normal or bimodal distributions most commonly appear in Method-B data, presumably because attenuation is a function of the variation in skill levels within the group of subjects at following the manufacturer's instructions, not the expertise in application or coaching of the experimenter. Figure A5 (Annex A) illustrates the variation in attenuation across subjects, including several devices (e.g., premolded earplug #1) that exhibit bimodality. WG11 has explored alternative ways of describing attenuation data, including bimodal and other distributions fit to the data, as well as direct computation of the percentile values for the attenuation at each OB frequency without reliance on fitting a model.

The direct calculation of the percentiles has appeal but also some shortcomings. First, there is no standardized and generally accepted formula for computation of percentiles. The comprehensive statistical analysis software from the SAS Institute (1995) offers several different percentile definitions. While the effect of these different definitions is very small when dealing with large sets of values, when computing the percentile for a small set of values (for example the ten or twenty attenuation values at one frequency for an earmuff or earplug respectively) the definition used can substantially change the answer. For this reason we did not pursue direct percentile computation of the octave-band APV.

However, computing percentiles on a large set of values is quite reasonable and is the basis for the NRP_A single-number rating and a companion NRP_G graphical rating we examined. NRP_A computes a large set (1000 or 2000) of protection values, one for each combination of subject and noise in the NIOSH 100, and directly finds the protection value corresponding to the desired protection rate target Q using the percentile formula described in Annex C. In the course of our work we found disadvantages to this approach as described in Section 3.2; this led us to the ratings we recommend, NRS_A and NRS_G. NRS_A computes the same large set of protection values as NRP_A but uses normal statistics to determine the protection corresponding to the targeted protection rate. Using normal statistics for the set of protections calculated for all subjects in a large set of noises offers two advantages over the traditional use of normal statistics applied to the attenuation data directly. First, it allows both across-spectrum as well as across-subject protection uncertainty to influence the high and low values in a two-number rating that describes the range of protection an HPD provides. Second, the across-spectrum uncertainty "smears" the modes that may be present in the attenuation data, creating a more uniform distribution of values. This is illustrated in Figure 7 where histograms of the protection for twenty subjects are shown computed in pink noise (left) versus the NIOSH 100 (right). The device chosen (premolded plug #1) exhibits one of the more non-normal distributions from the 20 HPD database (see Figure A5).

Figure 7 – Normality of protection value distribution improved when combined with many noise spectra.



3. RESULTS

In the previous section we described an extensive set of attenuation and noise data, proposed a number of new rating methods as well as prominent existing ratings to assess, and described a conceptual and computational framework by which to judge the ratings. In the following paragraphs we present the results of the calculations and the insights gleaned from them.

3.1 Rating values

Table 3 shows the values for all of the single-number ratings presented in this report for each of the 20 HPDs in the reference database of protectors. The protectors are grouped into four categories: earplugs (foam, premolded and semi-insert), earmuffs (band- or cap-mounted), moderate attenuators (one earmuff and one earplug with reduced high frequency and relatively flat attenuation) and one double protection combination (foam plug plus earmuff). Averages for each rating across all 20 HPDs are also shown. Multi-number ratings are shown graphically in Annex A.

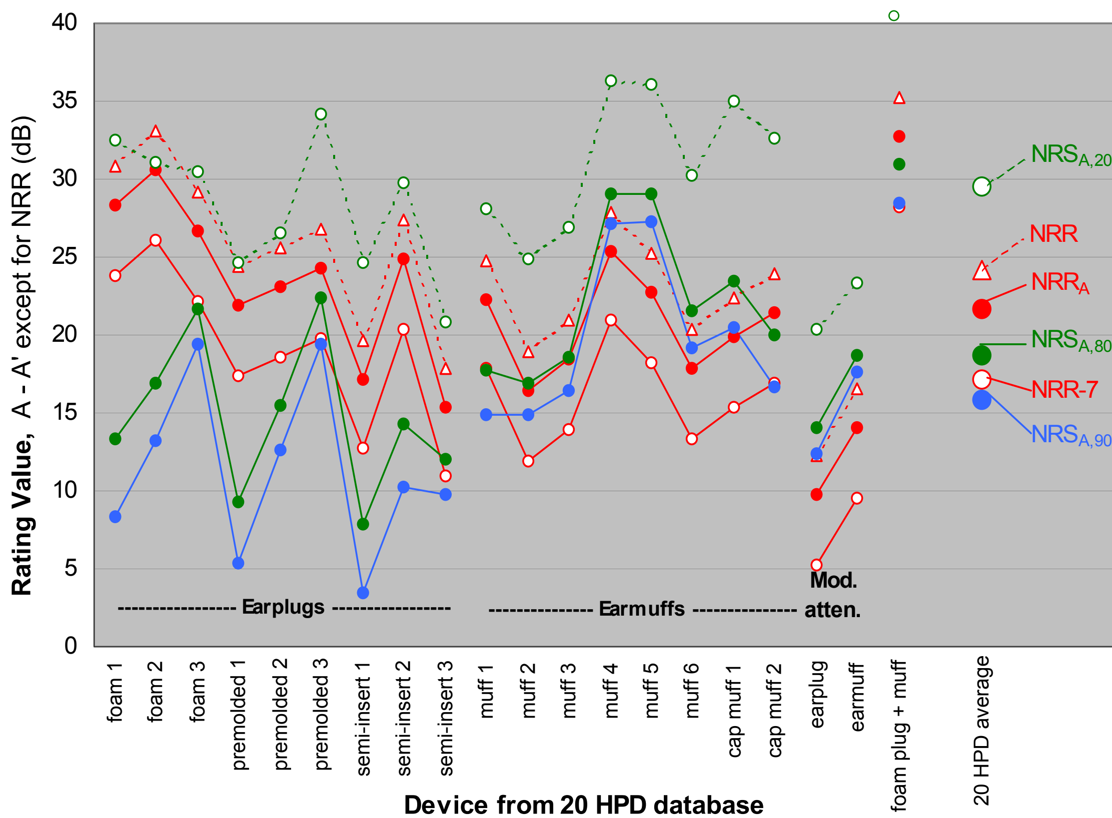
Rating values using attenuation data from Method-A and ISO 4869-1 tests are discussed in Section 3.6. Figure 8 plots the rating values across the 20 HPDs for the NRR and the new ratings we recommend. The specific protection targets for the low and high values in the two-number NRS_A range we recommend are *Q* = 80 and 20% for reasons discussed in the remainder of this section. We also show for consideration the 90th percentile NRS_A values and corresponding Waugh analysis metrics because this is an alternative, more conservative value that might be selected for the lower value in the range.

In the table and figure we show two ways of adjusting the NRR for use with A-weighted noise measurements, the –7 dB adjustment (NRR-7) recommended in current government guidelines and the –2.5 dB adjustment used earlier in the S3.19 bars shown in Figure 4. As results in the next section will illustrate, in industrial noise the –7 dB adjustment is overly conservative.

Table 3 – Rating values for the reference set of 20 HPDs. All ratings computed using Method-B data with the exception of the NRR and its variants that use the manufacturer's labeled S3.19 data. See Section 2.4 for definitions of the ratings. The recommended ratings are highlighted green.

			Earplugs									Earmuffs								Mod. atten.		
Type	Rating	**20 HPD average**	foam 1	foam 2	foam 3	premolded 1	premolded 2	premolded 3	semi-insert 1	semi-insert 2	semi-insert 3	muff 1	muff 2	muff 3	muff 4	muff 5	muff 6	cap muff 1	cap muff 2	earplug	earmuff	foam plug + muff
C – A'	NRR	**24**	31	33	29	24	26	27	20	27	18	25	19	21	28	25	20	22	24	12	17	35
	SNR_{84}	**21**	14	18	23	10	17	24	9	16	15	20	20	22	32	32	24	26	22	16	21	34
A – A'	NRR-7	**17**	24	26	22	17	19	20	13	20	11	18	12	14	21	18	13	15	17	5	10	28
	$NRR\text{-}7_{OSHA}$	**9**	12	13	11	9	9	10	6	10	5	9	6	7	10	9	7	8	8	3	5	14
	$NRR\text{-}7_{NIOSH}$	**6**	8	10	8	0	1	1	-1	1	-2	12	7	9	14	12	8	10	11	-3	5	19
	NRR_A	**22**	28	31	27	22	23	24	17	25	15	22	16	18	25	23	18	20	21	10	14	33
	Class	**16**	10	15	20	5	15	20	5	10	10	15	20	20	20	20	20	20	20	15	20	20
Low value, Q = 90%	$NRF_{A,90}$	**15**	7	12	18	4	12	18	2	9	10	15	15	17	28	27	19	20	16	11	16	28
	$NRS_{A,90}$	**16**	8	13	19	5	13	19	3	10	10	15	15	16	27	27	19	21	17	12	18	28
	$NRP_{A,90}$	**15**	6	11	19	4	13	20	3	5	9	14	15	16	27	28	19	20	16	12	17	28
Low value, Q = 84%	$NRF_{A,84}$	**17**	11	15	20	7	14	20	5	12	11	16	17	18	29	28	21	22	19	13	17	30
	$NRS_{A,84}$	**18**	12	16	21	8	14	21	6	13	11	17	16	18	28	28	21	22	19	14	18	30
	$NRP_{A,84}$	**17**	10	15	21	7	15	21	4	13	11	16	16	18	28	29	20	22	20	14	19	30
Low value, Q = 80%	$NRF_{A,80}$	**18**	12	16	21	8	15	22	7	14	12	17	17	19	29	29	21	23	20	13	18	31
	$NRS_{A,80}$	**19**	13	17	22	9	16	22	8	14	12	18	17	19	29	29	22	24	20	14	19	31
	$NRP_{A,80}$	**19**	13	18	22	8	16	22	5	16	13	19	17	18	29	29	22	24	21	15	19	31
High value, Q = 20%	$NRF_{A,20}$	**29**	32	31	31	25	26	34	25	30	20	26	23	24	34	34	28	34	31	20	22	41
	$NRS_{A,20}$	**29**	33	31	31	25	27	34	25	30	21	28	25	27	36	36	30	35	33	20	23	41
	$NRP_{A,20}$	**30**	31	32	31	25	27	36	26	30	21	28	25	27	36	36	31	34	32	20	23	41

Figure 8 – NRR and recommended rating values (NRS_A at Q = 90%, 80% and 20%) for the 20 HPDs.



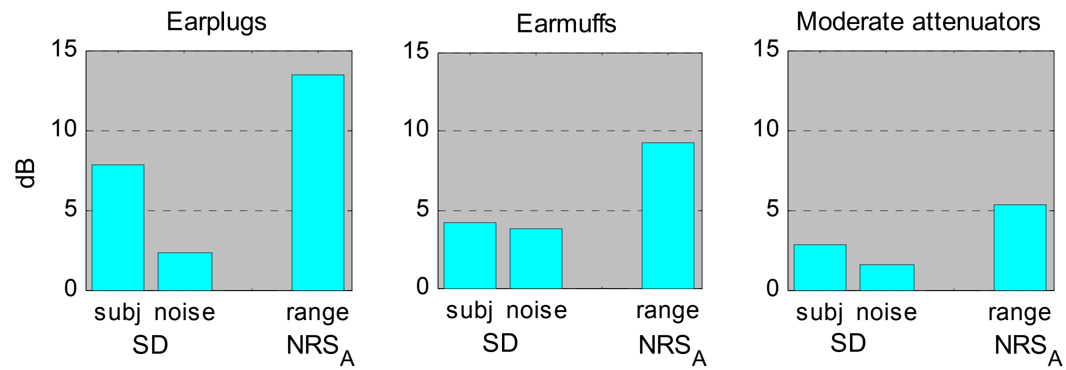
Examining the table and figure we make the following observations:

a) Comparing the value of NRR_A or NRR-7 to $NRS_{A,80}$ for plugs and muffs separately (Figure 8, solid red lines compared to solid green line) one sees that a new $NRS_{A,80}$ rating would lower the labeled performance of earplugs on average but increase the labeled performance of earmuffs. This reflects what has been observed in field studies (see Figure 4) and more correctly represents the performance of products than either the present OSHA- or NIOSH-recommended derating of the NRR shown in the table. The latter in particular leads to estimates of protection that are essentially zero for six of the nine earplugs in the 20-HPD database.

b) The ranges between the recommended low and high values (green highlighted rows, Table 3, and green lines, Figure 8) are notably smaller for the two moderate attenuation HPDs (5 dB on average) compared to earplugs (14 dB on average) and earmuffs (9 dB on average). This is the result of a rating method that factors both across-spectrum as well across-subject variation in protection into the design of the rating. Figure 9 shows these two sources of variation for each of these groups of devices. The figure shows that, for earplugs, across-subject variation dominates whereas for earmuffs the two sources of variation are approximately the same[10]. Both sources of

[10] The across-subject SD shown was computed for a single pink noise spectrum using the subject-mean OB attenuation for each subject. The across-noise SD was computed for the NIOSH 100 noises using the mean OB attenuation data for each HPD (essentially, APV_{50}). The SD values across all HPDs in each of the HPD sub-groups

Figure 9 – Variation in attenuation for different types of HPDs.



variation are less for the two moderate attenuators; this is reflected in the NRS_A range that indicates that the moderate attenuators are the more "predictable" HPDs.

c) The recommended lower rating value ($NRS_{A,80}$) averages to 19 dB across the 20 HPDs whereas NRR_A averages to 22 dB and NRR-7 averages to 17 dB. Thus, if the EPA were to adopt our recommendation of $NRS_{A,80}$ as the basis for a revised NRR rule, the values of ratings will actually increase slightly (2 dB = 19 – 17), on average across the HPDs in our database, when compared to current government guidance regarding how to apply the NRR to A-weighted noise. When compared to the more correct adjustment we use in the NRR_A (– 2.5 rather than –7 dB) ratings will decrease slightly (3 dB = 22 – 19). It must be noted, though, that these conclusions would change if a different selection of products were analyzed. We expect that using a larger proportion of earplugs more representative of the balance of products sold in the US would decrease the average $NRS_{A,80}$, bringing it closer to the average NRR-7.

3.2 Sensitivity of ratings to inter-subject variation

Method-B attenuation data exhibit larger across-subject standard deviations than Method-A, ISO-4869 or S3.19 data. In spite of this, one reason to base a revised NRR rule on Method-B data is that it should lead to better reproducibility of test results between laboratories (Murphy et al., 2004; Royster et al., 1996). This is due to three factors: (a) the removal of the influence of experimenter skill in HPD application or subject coaching, (b) elimination of the differences between laboratories in the skill level of an experienced, repeatedly-used subject pool that S3.19, Method A or ISO 4869-1 allow and (c) the higher standard deviations resulting from Method B, which broaden the confidence intervals used to statistically evaluate repeatability within laboratory and reproducibility between laboratories. Some have argued (Hall, 2003) that the larger standard deviation in Method-B data is indicative of a flawed test method that leads to a loss of precision in the results. To the contrary, we recommend Method B because the variability that it incorporates better correlates with available field studies of attenuation (Section 2.2, Figure 4) and because of the better repeatability of results between laboratories, particularly for reasons (a) and (b) above.

However, even though Method B is more repeatable than alternatives from a statistical perspective, manufacturers are understandably concerned about potential differences in rating values, in decibels not

---

shown in Figure 9 were computed after normalizing the A – A' values to be equivalent on average for each HPD. The NRS_A range is the average difference $NRS_{A,20}$ – $NRS_{A,80}$ for the group.

confidence intervals, which might arise from repeated testing of a product. The larger standard deviations in Method B data are a concern in this regard. The choice of rating method used and the underlying calculation and protection rate target can exacerbate this concern.

To examine this we use the last metric listed in Section 2.5, SD_S(Rating), the standard deviation of the rating value calculated across sets of subjects using the "bootstrap" method (Murphy, 2003). Table 4 shows the values computed using 4096 subject sets created by random resampling with repetition of the Method-B individual-subject data for each device[11].

Table 4 demonstrates that earplugs have a higher standard deviation of the rating value than earmuffs. This results from the larger SD in the underlying attenuation data. The Table also shows that the SD of the rating value increases when targeting 90% protection rather than 80%. What is most interesting is that NRP_A shows substantially more tendency for the rating value to vary with changing subject set than either NRS_A or NRF_A. Quantitatively, using NRF_A as the baseline for comparison since it uses the traditional method of computing single number ratings, at Q = 80% NRP_A shows 26% worse repeatability with changing subject sets than NRF_A for the 20 HPDs, and 39% worse repeatability for earplugs. By contrast, NRS_A compared to NRF_A shows a tiny, though not important, improvement in repeatability with changing subject sets. Comparing NRS_A at Q = 90% to Q = 80% one sees an increase of 15% in rating value standard deviation for both the 20 HPDs and for just earplugs.

Table 4 - Standard deviation of single-number rating value across multiple sets of subjects, SD_S(Rating), based on Method-B data with 4096 bootstrap iterations. The recommended rating is highlighted green.

Target protection rate	Rating	20 HPDs	Earplugs	Earmuffs
Q = 80%	NRF_A	1.9 dB	2.2 dB	1.8 dB
	NRS_A	1.9	2.2	1.7
	NRP_A	2.4	3.1	1.9
Q = 90%	NRF_A	2.3	2.6	2.2
	NRS_A	2.2	2.5	1.9
	NRP_A	2.7	3.2	2.3

These trends are not surprising. The standard deviation of the rating value across subject sets should be greater for data with more across-subject variability such as earplugs compared to earmuffs. It should also increase for higher values of the protection target Q because in that case the variability of the attenuation data factors more heavily into the rating computation (i.e., more SD are subtracted from the mean). That the NRP_A shows greater rating value SD across subject sets than either NRF_A or NRS_A is also not surprising. The latter two ratings are computed using normal statistics, where variability is measured relative to the mean. NRP_A on the other hand computes the rating directly using a percentile method, so the low-valued tail of the distribution of protection values drives this computation. Thus, NRP_A is inherently more sensitive to the influence of the lower performing subjects.

[11] With this many repetitions, repeated runs of the 4096-iteration calculation of the table (a four-hour process on a fast personal computer) showed repeatability of these values to better than 0.1 dB.

Note that SD_S(Rating) represents the true standard deviation of rating values that would be observed on many retests only if the distribution of attenuation across subjects in the data set used reflects the true distribution across the larger population of possible subjects. However, even if this assumption is not valid, these values represent the relative repeatability of different ratings and protection targets due to the underlying across-subject variability because the SD_S(Rating) is computed from the same attenuation data for each rating. If the aforementioned assumption is true[12] and repeated tests in different laboratories or within the same laboratory produce similar standard deviations in a rating value, then one can say that the 95% confidence range for repeatability (approximately ± 2 SD) is ±4 dB for NRS_A and NRF_A, while it is ±5 dB for NRP_A (20 HPDs, Q = 80% or 90%). For earplugs, the confidence range for NRF_A and NRS_A is ±4 dB (Q = 80%) and ±5 dB (Q = 90%) while it is ±6 dB for NRP_A at either protection rate target.

Because of the potential impact on manufacturers of rating values differing substantially upon retest, we strongly recommend NRS_A be adopted over NRP_A. This is also a factor in our recommending a protection rate target of 80% for the lower value in the two-number definition we propose for the primary label.

3.3 Waugh analysis using Method-B and industrial noise

Table 5 shows the values for the Waugh analysis metrics defined in Section 2.5. The metrics are shown for all ratings examined in this report, either at Q = 98% (–2 SD) for the NRR and its variants, computed from the manufacturer's labeled S3.19 data, or at Q = 84% (–1 SD) for the Method-B based ratings. We have chosen to show rating performance at Q = 84% in this table because it is the value that has been most commonly used with Method-B and ISO 4869-1 attenuation data. It is also approximately mid-way between the easier-to-relate-to targeted protection rates discussed in our goals for a rating in Section 2.4. Figure 10 shows histograms of the Waugh population A' values for most of the ratings shown in Table 5. Table 6 at the end of this section shows the Waugh analysis metrics for several of the ratings but at targeted protection rates of 80% and 90%.

Examining the upper section of Table 5 one can see that the protection rate ($P_{protected}$) for the NRR (61%) used as intended (NRR or NRR-7) significantly misses its 98% target. This is not surprising because S3.19 data are invariably higher than Method-B data; if ratings computed from it were taken as accurate as given on the label, HPDs could ostensibly be used in high levels of noise and many users (achieving actual protection values consistent with Method-B results) would be at risk of underprotection. In other words, the Waugh analysis shows that S3.19 data and the NRR can't be trusted to deliver protection rates close to the 98% target[13]. The NRR used as intended also results in much higher

[12] Though we do not report the details, it is worth noting that, in earlier work, the second author assembled a different selection of HPDs (15 earplugs, 4 earmuffs) and conducted Method-B tests in his laboratory and also contracted to have them tested in another laboratory. Computing ratings using the data from the two labs and computing the standard deviation of the rating differences for each HPD produced results quite similar to those shown in Table 4.

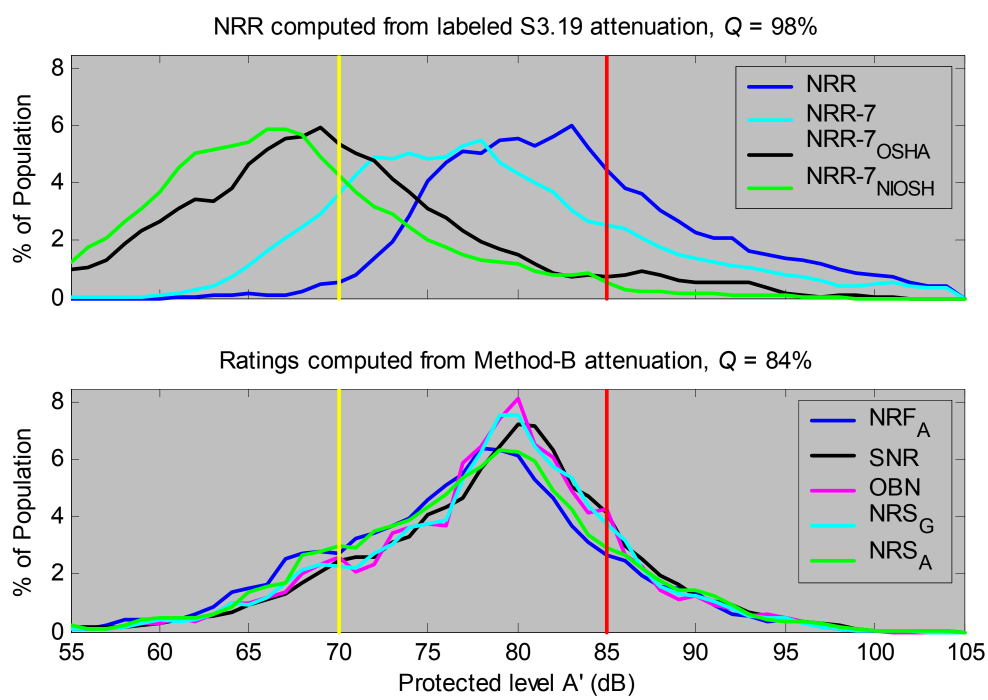
[13] Of course, if a Waugh analysis were done using S3.19 data for both the rating computation and to represent the Waugh population, the NRR would fare better. We have not done the Waugh analysis this way because there is no evidence that S3.19 data represent the performance achieved by groups of users in practice.

Table 5 – Waugh analysis metrics for all ratings using Method-B for the Waugh attenuation combining all 20 HPDs in the Ref300 noises. The ratings we recommend are highlighted in green.

	Rating	---------- Consolidated ---------- $P_{prot.}$ %	P_{ideal} %	A'_{under} dBA	$SD(A')$ dBA	Across noise spectra SD_N (P_{prot}) %	WC_N (P_{prot}) %	WC_N (A'_{under}) dBA	Across HPDs SD_H (P_{prot}) %	WC_H (P_{prot}) %	WC_H (A'_{under}) dBA
Rating attenuation = S3.19, target protection rate *Q* = 98%											
Present NRR	NRR	61	60	100	8.4	5	52	102	27	19	101
	NRR-7	78	66	98	8.7	7	55	98	18	42	98
	NRR-7_{OSHA}	93	40	91	8.4	3	88	93	8	76	89
	NRR-7_{NIOSH}	98	29	90	8.0	1	92	91	4	86	92
	NRR_A	63	60	100	8.7	12	30	101	24	25	101
Rating attenuation = Method B, target protection rate *Q* = 84%											
Single number A-A'	Class	85	62	91	8.3	8	59	93	10	56	89
	NRF_A	85	68	91	7.7	9	53	93	3	78	88
	NRS_A	83	69	91	7.6	10	49	93	2	76	89
	NRP_A	84	68	91	7.8	10	48	93	0.5	83	88
C - A'	SNR	81	70	91	7.3	4	74	91	3	74	89
Multi-number A-A'	NRS_G	83	70	90	7.2	3	75	90	4	74	88
	NRP_G	84	71	90	7.4	3	77	90	1	81	86
	OBN	83	71	90	7.1	1	81	90	5	73	88

Increasing *a priori* accuracy

Figure 10 – Histograms of Waugh populations for major ratings listed in Table 5. Boundaries for the ideal protection range are shown in red (underprotection) and yellow (overprotection).



equivalent level for the underprotected share of the population (A'_{under}), 7 to 10 dB higher than for the Method-B ratings. Figure 10 illustrates this – the NRR-based A' distributions are substantially broader than the distributions for the Method-B ratings. Only with derating (NRR-7_{OSHA} and NRR-7_{NIOSH}) does the underprotected equivalent level fall into line and the protection rate approach the target, but this is achieved at the price of a significant share of users potentially being overprotected, as can be seen from P_{ideal} values of 40% or less and in the figure, where one sees how derating shifts the broad NRR distributions to the left. Furthermore, use of OSHA's derating treats individual HPDs unfairly by assuming the same derating factor for all products; this is reflected in the high value of $SD_H(P_{protected})$ which measures how much the accuracy of a rating varies from HPD to HPD. NIOSH's derating by HPD type performs better in this regard, but is still unfair in that it represents the performance of all earplugs except formable foam ones as being essentially zero, much less than what Method-B-based single-number ratings indicate.

Turning to the lower section of Table 5, one can see that the protection rate is near the target of 84% for all of the ratings computed from Method-B data, with NRP_A and NRP_G hitting the target precisely as a result of their being based on a direct percentile calculation over a representative set of protection values. The greatest deviation from the target is the SNR at $P_{protected}$ = 81%. The various values for the ideal protection rate P_{ideal} and underprotected equivalent level A'_{under} are also quite similar for all of the ratings, and all of the distributions shown in the bottom half of Figure 10 look similar. That these ratings perform better than the NRR and its variants is not surprising since we test the rating (i.e., create the Waugh analysis A' population) using the same attenuation data from which the rating is calculated. This is valid as a means of testing which rating is most accurate at representing the performance of HPDs in a few simple-to-use numbers. But, to be as critical in examining the ratings we propose as we are with the NRR, in Section 3.7 we do a Waugh analysis to see how well Method-B achieves the targeted protection rate when using actual field data for the Waugh population.

How large a difference among the consolidated metrics is significant? The largest source of uncertainty that can affect the Waugh analysis is the set of attenuation data used. We can estimate the sensitivity to this factor by calculating how much the metric values change for expected changes in rating values resulting from a different set of subject data. If the NRS_A value decreases by 2 dB, which amounts to one across-subject-set SD [SD_S(Rating) = 2.0 dB at Q = 84%, all 20 HPDs; compare with Table 4] then $P_{protected}$ increases 6%, from 83% (see Table 5) to 89%. For the same NRS_A change, P_{ideal} decreases by 0.4% and A'_{under} decreases by 0.8 dBA (corresponding to a change of 17% in the noise dose).

Examining the histograms in the bottom portion of Figure 10 one can see that these observations make sense. As long as the peak of the distribution is reasonably well centered between the yellow overprotection and red underprotection fences then changes in the rating value that shift the distribution to the left or right will have much greater effect on $P_{protected}$ (the portion of the distribution to the left of the red underprotection fence) than P_{ideal} (the portion between the fences). Furthermore, the change in underprotected equivalent level should be somewhat less than half the change in rating value, as was calculated, with the actual value depending on the shape of the distribution's tail to the right of the red fence. For the other ratings similar sensitivities of metric value to changes in rating are to be expected because the histograms shown in Figure 10 are shaped similarly. So, considering separately the single-number and multi-number ratings shown in Table 5, ranges in variation of $P_{protected}$ (≤4%) and A'_{under} (≤1 dBA) are not significant or important as the differences are less than we compute for a 2-SD change in rating value. The only difference in the consolidated metrics in Table 5 that is significant is for the Class method; it's P_{ideal} value is 6 to 7% worse than the other single-number metrics because of its effective rounding of rating values to approximately 5-dB increments. This is also reflected in a larger *SD*(A').

The values of the consolidated metrics in Table 5 are so similar because the across-subject variability in the attenuation data dominates the calculation, largely overwhelming the differences in accuracy of the various ratings[14]. As a result the simplest, easiest-to-use rating (a number one simply subtracts from an A-weighted noise reading) can work as well as more complex methods, even the OB gold standard, when considered on average across a set of HPDs and industrial noise spectra. We must turn to other ways of looking at the Waugh A' population, such as the across-spectra and across-HPD metrics, to see the benefits of the more complex ratings systems as well as to determine which of the single-number systems to recommend. These are examined in the next two sections.

One of the rating goals described in Section 2.4 was to use an easy-to-relate-to targeted protection rate of either 80 or 90% to define the low value in the range. Table 6 lists Waugh analysis metrics for the single-number and multi-number ratings under consideration at these target values[15]. In general, the table shows that the Noise Reduction Statistics (NRS_A and NRS_G) produce the tightest A' distributions [smaller value of *SD*(A')] and best P_{ideal} values, though the differences just begin to become meaningful.

Table 6 – Waugh analysis metrics at preferred low-value protection rate targets of 80 and 90%. Method-B data used for both rating and Waugh population attenuation combining all 20 HPDs.

	Rating	Consolidated $P_{prot.}$ %	Consolidated P_{ideal} %	Consolidated A'_{under} dBA	Consolidated *SD*(A') dBA	Across noise spectra SD_N (P_{prot}) %	Across noise spectra WC_N (P_{prot}) %	Across noise spectra WC_N (A'_{under}) dBA	Across HPDs SD_H (P_{prot}) %	Across HPDs WC_H (P_{prot}) %	Across HPDs WC_H (A'_{under}) dBA
Single number ratings											
Q = 90%	NRF_A	89	66	90	8.0	8	60	92	2	84	88
	NRS_A	88	67	90	7.8	8	56	92	1	85	87
	NRP_A	90	65	89	8.3	8	61	91	0.4	89	87
Q = 80%	NRF_A	82	69	91	7.5	10	47	93	3	75	89
	NRS_A	80	69	92	7.5	11	46	94	3	74	90
	NRP_A	80	69	92	7.7	12	43	94	0.6	78	88
Multi-number ratings											
Q = 90%	NRS_G	88	71	89	7.5	2	83	89	3	81	86
	NRP_G	90	69	88	8.0	2	84	88	1	88	87
	OBN	90	73	89	7.5	1	88	90	4	81	88
Q = 80%	NRS_G	80	71	91	7.1	3	72	91	4	72	90
	NRP_G	80	70	91	7.2	3	73	93	2	76	87
	OBN	80	71	91	7.0	1	78	91	4	71	90

[14] If one compares Table 5 to the comparable metrics Waugh computed (Waugh, 1984), one sees higher *SD*(A'), the key metric on which Waugh focused, in our data. This is primarily because we used a database of HPDs with a much larger ratio of earplugs to earmuffs than Waugh did with the concomitant higher attenuation SD.

[15] Table 6 excludes the Class method because of its poor P_{ideal} performance, and the SNR for reasons discussed in the next section.

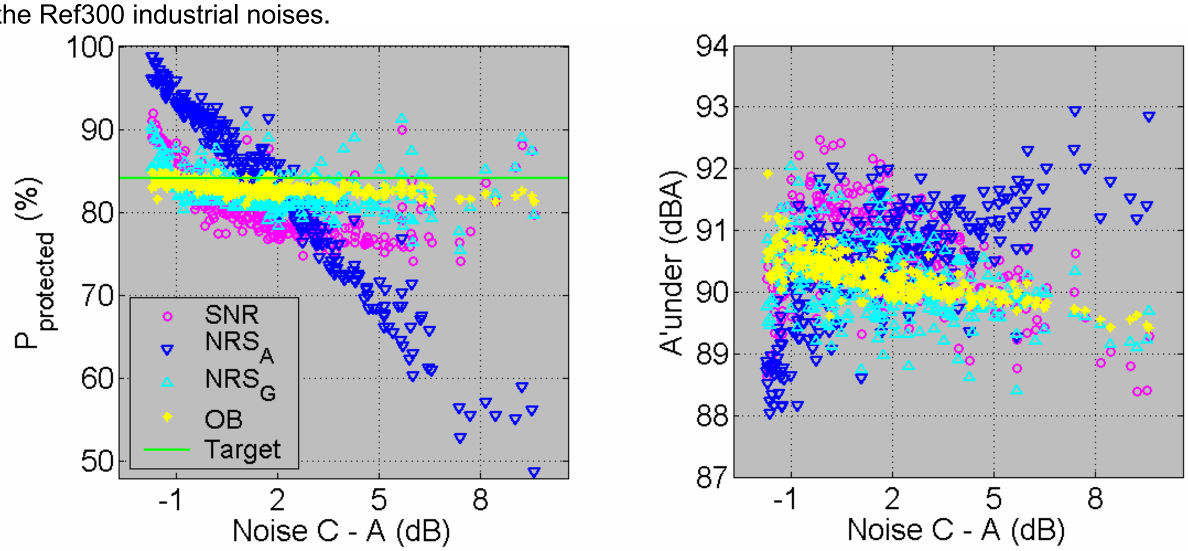
The Noise Reduction Percentile ratings show the widest A' distributions because, as discussed in Section 3.2, they are most sensitive to outliers in the attenuation data. Also, comparing protection rate targets of 90% to 80% one sees that the former, more conservative, target decreases the equivalent level for the underprotected share of the population by 1 to 3 dB (1.7 dB for NRS_A) and the ideally protected rate P_{ideal} changes little (i.e., the overprotection rate increases).

3.4 Accuracy of ratings for different noise spectra

The middle set of columns in Table 5 show how $P_{protected}$ varies across the set of noise spectra used in the analysis, the Ref300 set of industrial noises. The first of these columns (SD_N) is the standard deviation of $P_{protected}$ while the second and third columns (WC_N) give the worst-case (lowest) value of $P_{protected}$ observed for any of the spectra and the highest underprotected equivalent level coinciding with that worst-case value of $P_{protected}$. The SD_N values show an advantage for a C – A' rating like the SNR; it results in less than half the protection-rate variation across industrial spectra seen with the A – A' ratings and notably better worst-case protection rates. However, in situations where C-weighted noise data are available, applying it through looking up the protection corresponding to the noise's C – A in either of the graphical ratings (NRS_G or NRP_G) achieves somewhat better values for these metrics. OBN, the "gold standard" achieves the smallest variation in $P_{protected}$ across noise spectrum since it is based on the greatest amount of information about the noise spectrum.

Figure 11(left panel) illustrates the across-spectrum performance of the different types of ratings (A – A', C – A', graphical, and OB) in an alternative manner by plotting $P_{protected}$ as a function of C – A value for the noise spectrum. Each dot on the graph represents the protection rate achieved by 310 subjects[16] in one of the spectra in the Ref300 database. This chart ignores the individual variability across subjects, rather characterizing the performance for a group of wearers with a single number for each spectrum. Viewed in this way, the change in protectiveness of an A – A' rating with changing C – A values is apparent.

Figure 11 – Variation with noise spectrum of the protection rate (left panel) and underprotected equivalent level (right panel) for different rating types, and all 20 HPDs. Each mark is one spectrum in the Ref300 industrial noises.



[16] 20 subjects for each of the eleven earplug sets of attenuation data and 10 subjects for the nine earmuff sets of attenuation data in the 20-HPD database.

However, this method of presentation may exaggerate the effect since changes in the $P_{protected}$ can simply arise due to subjects just sliding over the fence by portions of a decibel. Figure 11(right panel) illustrates this effect by showing that the change viewed in terms of the underprotected equivalent level is only about 1.5 dB for a C – A = 6 dB (compare the NRS_A values at about 91.5 dBA to those for the OB method at about 90 dBA).

Figure 12 similarly portrays the variation in protection rate with noise C – A value for the same ratings, but this time the 20 HPD database is separated into earplugs (9 products), earmuffs (8 products), and flat/moderate attenuators (2 products, one plug and one muff). We also show a graph for a truly flat attenuator, the Etymōtic Research ER-15 custom molded earplug[17]. We chose the latter as an example of what many consider an “ideal” attenuator because it minimally distorts the sound heard by the wearer, simply reducing the level. Figure 12(a) shows the shape of the attenuation response (the average APV_{84}) for the four categories of devices depicted in the other subplots. Finally, note that each mark in Figure 12 represents the protection rate in one of the 70 AF+CA noise spectra described in Annex B. For this chart we use these noises instead of the Ref300 because they are a smaller set of spectra, resulting in a less crowded graph, with widely distributed C – A values.

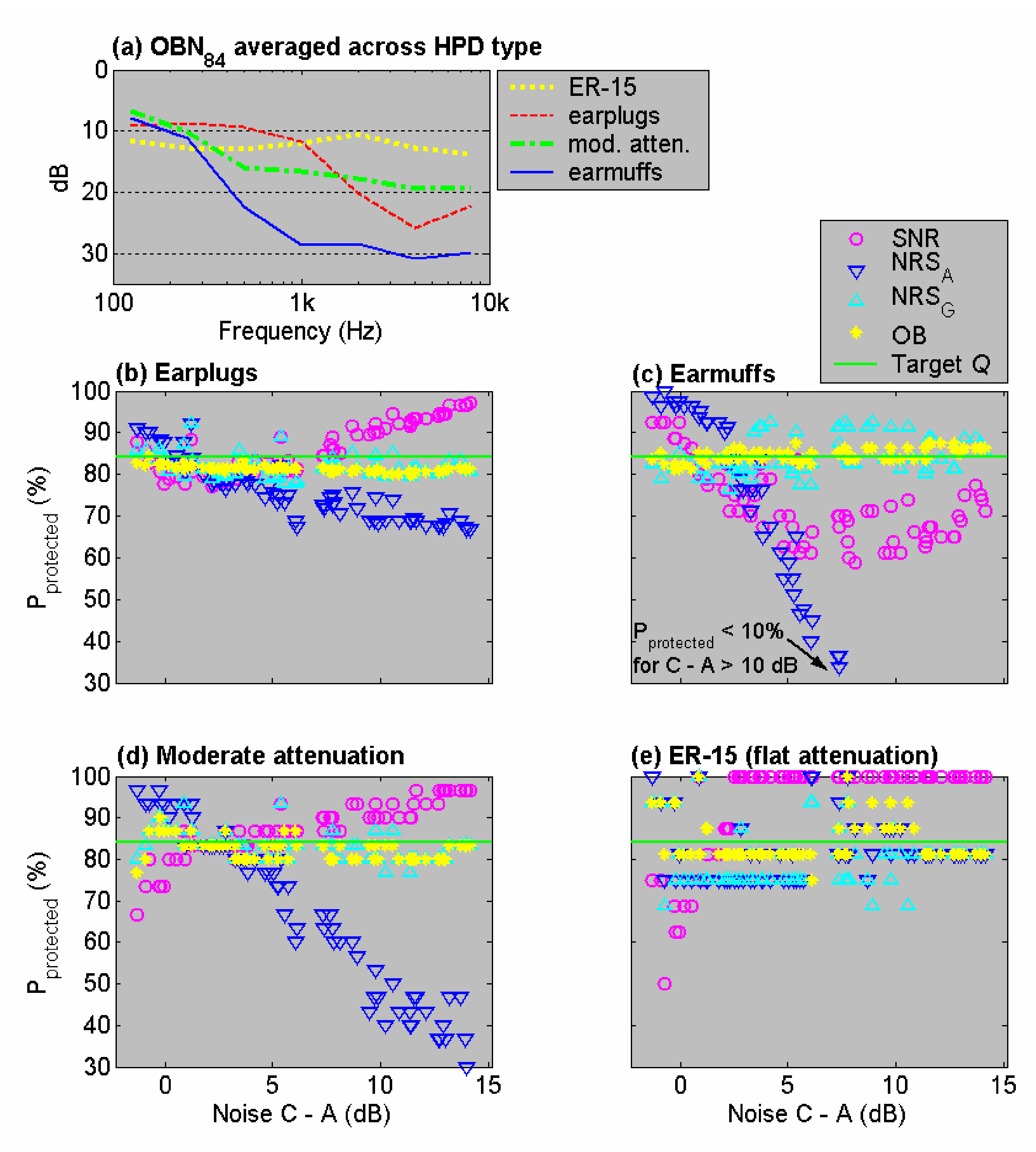
Figure 12 shows that, over a wide range of noise spectra, an A – A’ rating like NRS_A can lead to significant deviations from the protection rate target. How large the error is depends on the attenuation response (i.e., variation with frequency), with earmuff-like attenuations resulting in the greatest errors. If the figure showed NRF_A it would perform similarly to NRS_A; in fact, Table 5 shows incrementally better across-spectra performance for the traditionally computed NRF_A versus the new and proposed NRS_A for the low C – A values that dominate the Ref300 noises. Recall that NRF_A is calculated from the APV in pink noise and then a constant adjustment is subtracted so as to achieve 84% protection in industrial noise. The correct value of that adjustment is a function of the range of noise C – A value assumed and the average shape of the attenuation response. The rather dissimilar distributions of dark blue marks for the different HPD types shown in Figure 12 indicate that the 3-dB adjustment that is correct on average for the 20 HPDs is a compromise. The advantage of the NRS_A is that no such adjustment (or compromise) is needed; the calculation of the rating using the NIOSH 100 noises in essence self-calibrates the rating value, for the given HPD, to come as close to the protection target as it is possible for an A – A’ rating based on normal statistics to achieve. This allows the NRS_A to perform more consistently for different protectors, including even flat attenuation devices like the ER-15 (see the next section).

Figure 12 also shows that C – A’ ratings like the SNR perform better than A – A’ ratings do though, as noted earlier, the NRS_G performs even better using no more information about the noise. The SNR also exhibits deviations from the target protection rate that are dependent on the shape of the protector’s attenuation response. The SNR performs best for earplug-like attenuation responses and the error, for spectra dominated by low frequency energy (high C – A), is about the same for the SNR as for the NRS_A though in the opposite direction. For the flat-attenuating ER-15, the SNR shows the largest error and this occurs at low C – A values. In contrast, NRS_G consistently results in protection rates near the target for all HPD types and C – A values. Because of this divergent performance of C – A’ ratings for different types of HPDs, as well as because of the potential confusion and error regarding whether a single-

---

[17] Since no Method-B attenuation data were currently available for the ER-15, values in accordance with S12.6-1984, a procedure similar to Method-A were used. This yields smaller across-subject variation than the Method-B data used elsewhere in Figure 10.

Figure 12 – Variation in protection rate across noise spectra (C – A) for different types of HPDs and ratings. Computed in 70 aviation noises with Q = 84% using Method-B for rating and Waugh attenuation. The ER-15 shows greater $P_{protected}$ variation for even the better ratings, as compared to the other device types, because the protection distribution for this device is so small [SD(A') = 1.4 dB for OBN_{84}; compare with Table 5]. This makes $P_{protected}$ for this device sensitive to slight errors in a rating's estimation of A'.



number rating is to be subtracted from A-weighted or C-weighted noise measures, we reject the SNR from further consideration[18].

OBN shows the best across-spectrum performance in Table 5 and the tightest clustering of points near the target lines in Figures 11 and 12. However, this accuracy comes at the expense of complexity of both data collection (octave-band spectra instead of broad-band measures from a sound level meter or dosimeter) and calculation (computation using logarithms instead of simply reading from a graph or, at most, interpolating a table). In fact, given the $\pm$ 2.5-dBA accuracy tolerance on noise measurements (see Section 1.1) and that a 2.5-dBA change in noise level changes $P_{protected}$ by approximately 8%[19], the improvement in $SD_N(P_{protected})$ from 3% to 1% in going from NRS_G to the more complex OBN is not important.

Table 7 shows the various Waugh analysis metrics computed in the seventy aviation noises (AF+CA) to examine how well various ratings perform in spectra rather different from the industrial noise environments for which they are primarily intended. The table demonstrates how poorly an A – A' rating does at achieving consistent protection rates for widely varying noise spectra. However, it again shows that NRS_G performs virtually as well as the more complex OB method with much greater ease of use. The table also shows that using NRS_G instead of NRS_A increases P_{ideal} a significant amount (57% to 69%) and decreases the underprotected equivalent level by 2 dBA. For the worst-case spectrum, changing from NRS_A to the graphical C – A rating reduces A'_{under} by 5 dBA. Thus, an NRS_A-based labeling rule should include guidance describing under what conditions using the graphical NRS_G is advised.

Table 7 – Waugh analysis metrics in non-industrial noise at protection rate target Q = 84%. Computed for all 20 HPDs consolidated using Method-B for rating and Waugh attenuation in the combination of the AF50 and CA20 noise databases.

		——— Consolidated ———				Across noise spectra			Across HPDs		
	Rating	$P_{prot.}$ %	P_{ideal} %	A'_{under} dBA	$SD(A')$ dBA	SD_N (P_{prot}) %	WC_N (P_{prot}) %	WC_N (A'_{under}) dBA	SD_H (P_{prot}) %	WC_H (P_{prot}) %	WC_H (A'_{under}) dBA
Single number A - A'	NRF_A	70	58	92	8.5	15	48	94	16	38	93
	NRS_A	67	57	92	8.3	15	45	95	14	40	93
C - A'	SNR	81	65	90	8.2	4	75	91	9	55	87
Multi-number A - A'	NRS_G	83	69	90	7.5	2	78	90	4	74	88
	OBN	82	69	90	7.5	1	81	89	5	73	88

[18] For a truly flat attenuator, A – A' would be a constant so the SNR understates the device's protection by 1 dB for every 1-dB increase in noise C – A value. In high C – A spectra this could result in significant overprotection. Figure 12(e) shows $P_{protected}$ quickly reaching its maximum possible value of 100%; P_{ideal} would continue to decrease as C – A increases.

[19] Calculated for OBN_{84} in the Ref300 noise spectra combining all 20 HPDs.

3.5 Accuracy of ratings for different HPDs

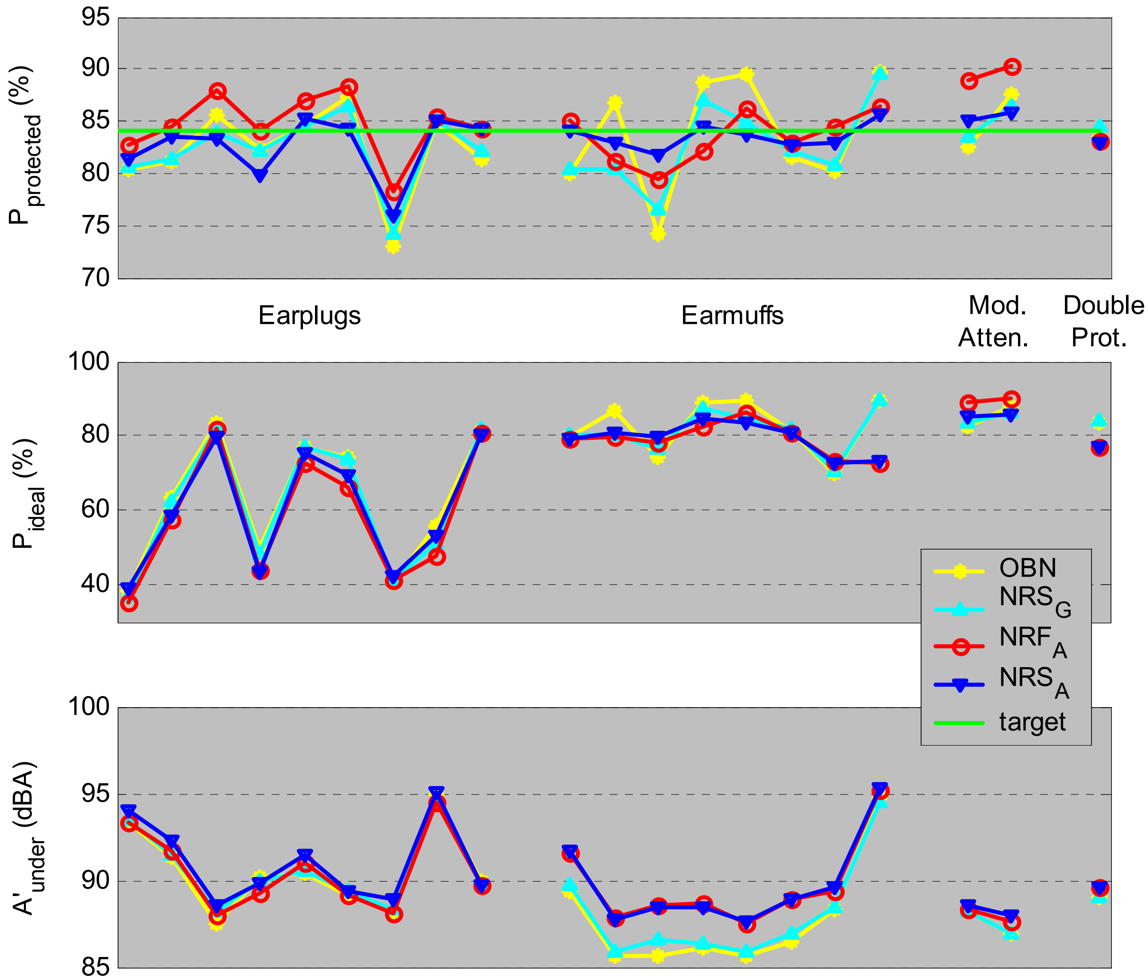
One of the goals for a new rating we described in Section 2.4 was that it should work consistently for all HPDs to which it was applied. The across-HPD columns in Table 5 address this issue. They show that, because of the 5-dB rounding inherent in the Class method, the standard deviation of protection rate across the twenty devices in the database is substantially larger than for the other single-number ratings and the worst-case $P_{protected}$ significantly misses the target. Among the other single-number ratings, only NRP_A stands out as more consistently achieving the targeted protection rate across HPDs; this is because of its direct computation of the percentile value. This benefit of NRP_A is not sufficient in our opinion to offset the increased sensitivity to changing subject sets shown in Table 4. It is also interesting to note in Table 5 that OBN exhibits higher $P_{protected}$ variation across HPDs than the single-number ratings. We suggest this is because, as a rating becomes more accurate in different spectra, it becomes more sensitive to differences in attenuation response for different HPDs.

Table 8 examines the consistency of performance across HPDs another way, by showing the consolidated Waugh metrics (across noise and HPD) separately for earplugs, earmuffs, and the moderate attenuation devices in the database. Figure 13 plots the values of these metrics for each HPD in the database for several of the rating methods. Table 8 shows that NRS_A more consistently achieves the targeted protection rate of 84% for each type of HPD than NRF_A. This is also seen in the upper plot in Figure 13 where the blue NRS_A line lies closer to the green target line than the alternative NRF_A rating of the same type and in the smaller value of $SD_H(P_{prot.})$ for NRS_A in Table 5. The effect is small and appears unimportant. The difference between NRS_A and NRF_A is more notable for a truly flat attenuator like the ER-15 discussed in the previous section. For the ER-15, the $NRF_{A,84}$ value is 10 dB whereas the $NRS_{A,84}$ value is 12.3 dB, identical (to within 0.1 dB) to the average across frequencies of the nearly flat attenuation [yellow line, Figure 12(a)]. The constant adjustment built into the NRF_A causes it to understate the performance of this exemplary device by 2 dB. This error in the ER-15's rating by $NRF_{A,84}$ results in a virtually 100% protection rate whereas $NRS_{A,84}$ results in an 80% protection rate, closer to the targeted value. This is the result of the absence of any assumptions about the frequency dependence of attenuation being built into the NRS_A as it is in the NRF_A and other existing single-number ratings.

Table 8 – Consolidated Waugh analysis metrics for different types of HPDs. Method-B used for both rating computation (Q = 84%) and Waugh population.

	Rating	Earplugs (9 HPDs)				Earmuffs (8 HPDs)				Moderate atten. (2 HPDs)			
		$P_{prot.}$ %	P_{ideal} %	A'_{under} dBA	$SD(A')$ dBA	$P_{prot.}$ %	P_{ideal} %	A'_{under} dBA	$SD(A')$ dBA	$P_{prot.}$ %	P_{ideal} %	A'_{under} dBA	$SD(A')$ dBA
Single number A-A'	NRF_A	85	59	91	8.6	84	79	90	5.9	89	89	88	3.4
	NRS_A	83	60	92	8.6	84	79	90	5.8	85	85	88	3.4
	NRP_A	84	60	91	8.8	84	80	91	5.8	84	84	88	3.4
C - A'	SNR	81	63	92	8.4	80	79	89	4.7	81	81	88	3.2
Multi-number A-A'	NRS_G	82	62	91	8.3	83	82	89	4.5	84	84	88	3.1
	NRP_G	84	63	91	8.5	84	83	89	4.5	85	85	88	3.1
	OBN	82	63	91	8.3	84	83	89	4.4	84	84	88	3.1

Figure 13 – Metric values for individual HPDs using same conditions as in Table 8.



3.6 Comparison of Method B to Method A and ISO 4869

In Section 2.2 we discussed the issue of whether a new rating should be based on Method-B attenuation data or the alternative experimenter-supervised fit procedures given in Method A and ISO 4869. In this section we examine through a Waugh analysis the protection rates that would be achieved if either of the latter two procedures were adopted as the basis for the rating. As we did when applying the Waugh analysis to the NRR, in this section we continue to use Method-B data to model the Waugh population because it has been shown to correlate best with field attenuation data (see Section 2.2).

We base our analysis of Method A on data provided to the authors by John Hall of the Air Force Research Laboratory Bio-Acoustics group at Wright-Patterson AFB (Hall, 2004). AFRL measured five HPDs (two foam earplugs, three earmuffs) following both a Method-B and a Method-A protocol. We base our analysis of ISO 4869 on data measured to that standard in various laboratories in Europe; the products included are the subset of the 20 HPDs for which individual-subject ISO-4869 data were available (see Figure A6). Table 9 gives the values for the recommended single-number rating ($NRS_{A,80}$) for these two HPD groups using Method-B data as well as data from the two experimenter-supervised-fit

Table 9 – Comparison of NRS_A values computed from Method-B, Method-A and ISO 4869 data.

	AFRL Method-A and B data for 5 HPDs							Subset of 20-HPD database with available ISO 4869 individual subject data												
			Plugs		Muffs					Earplugs					Earmuffs			Mod. atten.		
Rating	Method	**average**	foam A	foam B	muff A	muff B	muff C	Method	**average**	foam 2	foam 3	premold 3	semi-ins 1	semi-ins 2	muff 2	muff 3	muff 6	earplug	earmuff	foam plug + muff
$NRS_{A,80}$	B	**21**	17	19	23	20	28	B	**19**	17	22	23	8	14	17	19	22	14	19	31
	A	**28**	30	28	26	25	29	ISO	**26**	29	26	37	20	25	20	21	24	19	20	41
$NRS_{A,99}$	A	**21**	23	20	20	18	23	ISO	**19**	22	20	32	9	18	13	14	17	15	17	35

procedures (Method A and ISO 4869); an additional protection rate target ($NRS_{A,99}$ computed using a 2.3-SD correction) is also shown.

Table 9 shows that both experimenter-supervised fit procedures lead to ratings that exceed Method-B values, with the differences being largest for earplugs. This is consistent with the results for ISO 4869 presented in Figure 5. On average across all devices, the table shows that the NRS_A values calculated from data obtained using either of the experimenter-supervised-fit procedures exceed the values calculated from Method-B data by 7 dB. Similarly, values computed from either the Method-A or ISO-4869 data require a larger SD adjustment [2.3-SD (Q = 99%) for A/ISO vs. 0.84 SD (Q = 80%) for Method B], so that the Method A or 4869 values match ones computed from Method B (compare $NRS_{A,80}$ for Method B and $NRS_{A,99}$ for Method A). The congruence of the difference in average rating values and the additional SD adjustment needed with A/ISO data to obtain the same average rating for both sets of data, lends credence to the assertion that Method A and ISO 4869 yield largely similar attenuation results. This conclusion must be tempered by the limited Method-A data we have available: five devices and only two plugs, with the latter being more sensitive to procedural differences. Nevertheless, in the balance of this section we assume the procedures are indistinguishable, pool the values from Method A and ISO 4869, and refer to such data as Method A. This creates a collection of sixteen HPDs, eight of which are earplugs, seven are earmuffs and one is a muff/plug combination — a mix of devices similar to the 20 HPD database used in the foregoing sections.

Some members of WG11 have suggested that ratings be based on Method-A data because it reduces the influence of subject skill, and that the resulting ratings be derated in some way to produce values that approximate those seen in the real world. This derating of better-than-real-world attenuation data could be built into the rating computation itself; three possible approaches for derating NRS_A and NRS_G ratings computed from Method-A values are:

a) use a larger SD adjustment than is used with Method-B data,
b) subtract a constant correction in decibels from the rating or
c) multiply the rating by a constant.

These approaches can also be combined, e.g. a larger SD adjustment and a subtractive constant. Given the good correlation of Method-B ratings to values computed from field data and our recommendation of protection rate targets of Q = 80% and 20% for the low and high values of a Method-B rating, we propose that the goal for a Method-A rating should be to provide the same degree of protection. In other words, a Waugh analysis using the rating computed using Method-A data and derated in some fashion should yield protection rates of 80% using the low rating value and 20% using the high rating value when Method-B data are used for the Waugh population.

Turning first to the additional SD adjustment approach, the question arises as to how many SD to use. Table 9 shows that 2.3 SD ($NRS_{A,99}$) yields average rating values equal to those computed from Method B. However, Waugh observed (Waugh, 1984, Figure 6) that increasing the number of SD subtracted from the mean increases the value of *SD*(A') indicating broadening of the protected level distribution. We have confirmed this result when using the same Method-B data for both the rating and Waugh population. When we repeated this calculation using Method-A data for the rating and Method-B data for the Waugh population using the 16 devices listed in Table 9, we found that *SD*(A') did not monotonically increase with the number of SD used but rather had a minimum at a 2-SD adjustment. This occurs because the correlation across the 16 HPDs between the Method-A and Method-B rating values increases with the number of SD used, offsetting the effect observed by Waugh. Since approximately a 2-SD adjustment yields the tightest distribution of Waugh population protected levels, we choose to use that value and apply a small subtractive constant to further adjust the rating value to achieve the protection rate targets. A larger SD adjustment also increases the sensitivity of the rating value to changes in attenuation data during retesting, reflected in a larger SD_S(Rating) — this also argues for minimizing the number of SD used.

Regarding the second and third approach, we choose to use the same SD adjustment as is used in our recommended Method-B rating (±0.84 SD, Q = 20% and 80%). We also propose that a multiplicative constant is preferable to a subtractive one because the latter might lead to confusing negative rating values for the poorer performing HPDs.

Table 10 shows the results of a Waugh analysis using these two Method-A rating definitions compared to the Method-B ratings. Table 10 shows that a small subtractive correction of 3 dB must be applied when using a –2-SD adjustment in order to achieve approximately the 80% protection rate goal; the same correction applies to both NRS_A and NRS_G. For the high value, using a +2-SD adjustment requires a larger 8-dB correction to achieve approximately the 20% protection rate goal. When using the ±0.84-SD adjustments on Method-A data the required multiplicative corrections are 0.71 and 0.88 to achieve the low and high protection rate targets respectively.

Upon examination Table 10 shows that, without derating and if taken at face value, a rating definition based on Method A would lead to very poor protection rates (less than 40%) and high underprotected equivalent levels (98 dBA, 7-dB worse than for a Method-B rating). Of the two derating approaches, the 0.84-SD / multiplier approach offers markedly better SD of the rating value across subject sets than the 2-SD / subtracted-constant approach (0.8 vs. 1.5 dB) whereas the latter is virtually no different than the Method-B rating. This is not surprising. The lower SD_S(Rating) value for Method-A $NRS_{A,\,-0.84SD}$ (1.1 dB) compared to Method-B $NRS_{A,\,-0.84SD}$ (1.6 dB) reflects the lower across-subject variability of the former procedure; using a larger SD adjustment on the Method-A data amplifies the effect of across-subject variability, offsetting Method A's advantage. Conversely, derating via a multiplier further reduces the variability of rating values across subject sets — SD_S(Rating) decreases from 1.1 dB to 0.8 dB. Combining the effect of lower variability Method-A data, at the same SD adjustment, and a multiplicative derating factor results in an SD_S(Rating) that is half the value computed for Method B. Since a key argument for using Method A is its increased precision (repeatability of rating value on retesting) this argues strongly for the 0.84-SD / multiplier rating definition over the alternative 2-SD / subtracted-constant approach.

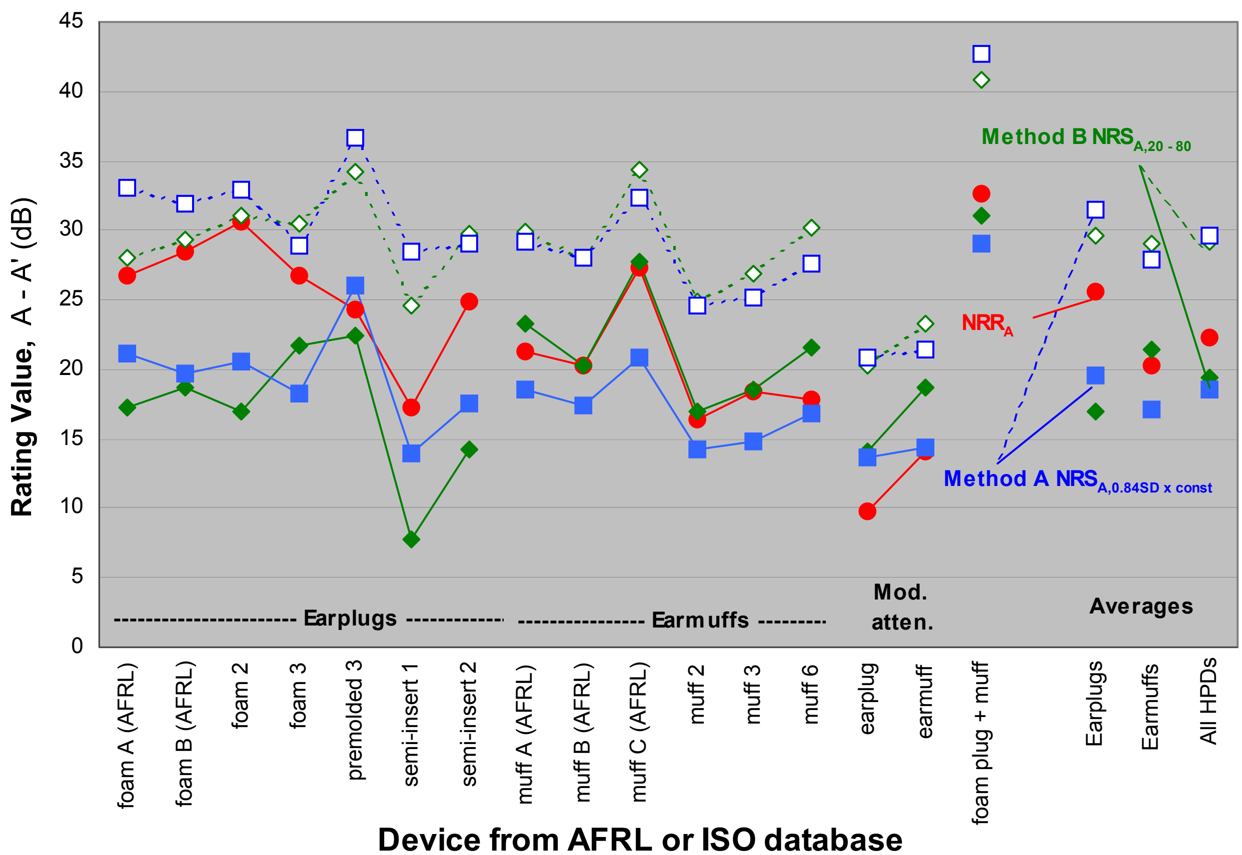
Table 10 – Waugh analysis of NRS_A and NRS_G ratings computed from Method-A / ISO-4869 data using various derating approaches and compared to Method B. Method-B data used for the Waugh attenuation in all cases; noise spectra are the Ref300.

	Rating	---------- Consolidated ---------- $P_{prot.}$ %	P_{ideal} %	A'_{under} dBA	$SD(A')$ dBA	Across noise spectra SD_N (P_{prot}) %	WC_N (P_{prot}) %	WC_N (A'_{under}) dBA	Across HPDs SD_H (P_{prot}) %	WC_H (P_{prot}) %	WC_H (A'_{under}) dBA	SD_S (Rating) dB
Rating attenuation = Method B												
	$NRS_{A,\ -0.84SD}$	80	72	91	6.8	13	42	93	3	74	90	1.6
	$NRS_{A,\ +0.84SD}$	21										
	$NRS_{G,\ -0.84SD}$	80	74	91	6.3	3	72	90	5	72	90	
	$NRS_{G,\ +0.84SD}$	20										
Rating attenuation = Method A / ISO 4869												
No derating	$NRS_{A,\ -0.84SD}$	39	39	98	7.3	17	11	100	20	11	101	1.1
	$NRS_{G,\ -0.84SD}$	33	33	97	6.6	5	28	99	18	8	101	
Adjust by 2 SD & constant	$NRS_{A,\ -2SD,\ -3dB}$	80	71	94	7.4	8	50	94	16	42	94	1.5
	$NRS_{A,\ +2SD,\ -8dB}$	17										
	$NRS_{G,\ -2SD,\ -3dB}$	79	74	93	6.6	4	72	93	16	39	94	
	$NRS_{G,\ +2SD,\ -8dB}$	25										
Adjust by 0.84 SD & multiplier	$NRS_{A,\ -0.84SD,\ x0.71}$	80	74	94	7.0	10	46	94	14	59	95	0.8
	Plugs only	73	67	95								
	Muffs only	95	89	87								
	$NRS_{A,\ +0.84SD,\ x0.88}$	20										
	$NRS_{G,\ -0.84SD,\ x0.71}$	81	78	94	6.5	6	68	94	15	59	94	
	$NRS_{G,\ +0.84SD,\ x0.88}$	22										

Table 10 also shows that while Method A, with the multiplicative derating factors shown, and Method B are comparably accurate in achieving the protection rate target, they differ in other regards. Compared to Method B, the derated Method-A rating has a 3-dB worse underprotected equivalent level (a 100% noise dose increase) and substantially worse across-HPD performance [larger $SD_H(P_{prot.})$ with an additional 13 to 15% of the population underprotected at an equivalent level that is 4 to 5 dB worse for the worst-case HPD]. The latter is not surprising as the correlation between the Method-B data used for the Waugh population and ISO 4869 (and presumably Method A) is not good (see Figure 5); this means that the derating under-corrects some devices while over-correcting others. This can also be seen in the last two rows of the table which show that the multiplicative derating chosen based on the sixteen HPDs in aggregate is not sufficient for earplugs (leading to only a 73% protection rate) while it is more than is needed for earmuffs (95% protection rate).

Figure 14 compares, device by device, the values of the NRR, our recommended rating (Method B, $NRS_{A,\ \pm0.84SD}$) and the more precise alternative (Method A, $NRS_{A,\ -0.84SD,\ x0.71}$ and $NRS_{A,\ +0.84SD,\ x0.88}$) along with averages across the sixteen devices and for earplugs and earmuffs separately. Note that Method B and the adjusted Method-A values track reasonably well except that, as with the NRR_A, the Method-A data rate average earplug performance as better than average earmuff performance, albeit by a smaller amount (3 dB using the derated Method A vs. 6 dB with the NRR

Figure 14 – NRR_A and Method-A / B rating values (NRS_A at Q = 80% and 20% from Method-B data; NRS_A with –0.84 SD and 0.88 or 0.71 multiplier for Method A) for devices listed in Table 9.



computed from S3.19 data). Method B, on the other hand, provides answers more reflective of field experience with earmuffs, on the average, outperforming earplugs by 4 dB.

To improve the accuracy for different HPD types, one could of course use HPD type-specific derating factors but these would only be valid for the set of devices used to determine them. In fact the conclusion that a Method-A rating is apparently as accurate as Method B, at least in aggregate across all devices, must be qualified by the fact that the derating factor is fit to this set of devices. Method A is unlikely to perform as well for a different set of devices and attenuation data. Fundamentally, the only way to insure accuracy of the rating is to use attenuation data that correlates well with real-world use of the product, which Method B accomplishes to a greater degree than Method A (see Section 2.2). Before a derated Method-A rating is adopted, at a minimum an interlaboratory study should be performed on a large and representative set of devices, using both Method B and either the Method A or ISO 4869 protocols (see Section 4.6). Only with such data could the derating factor be determined with confidence.

Some might criticize the foregoing comparison of Method B and Method A because the Method-B calculations are "circular" — the same Method-B attenuation data that are used to calculate the rating value are also used to model the Waugh population. To address this concern, Table 11 shows Waugh analyses for a subset of four earplugs using our recommended Method-B ratings, NRS_A and NRS_G (low

Table 11 – Waugh analysis comparing the Method-B and Method-A ratings using different laboratories for the rating and Waugh attenuation data. E•A•RCAL data are used for the Waugh population in all cases. The HPDs used are four similar-design earplugs, three foam and one premolded.

		---------- Consolidated ----------				Across noise spectra			Across HPDs		
Rating Method	Rating	$P_{prot.}$ %	P_{ideal} %	A'_{under} dBA	$SD(A')$ dBA	SD_N (P_{prot}) %	WC_N (P_{prot}) %	WC_N (A'_{under}) dBA	SD_H (P_{prot}) %	WC_H (P_{prot}) %	WC_H (A'_{under}) dBA
B (VT)	$NRS_{A,\ -0.84SD}$	83	68	93	7.8	6	69	94	10	73	94
A/ISO	$NRS_{A,\ -0.84SD,\ x0.71}$	74	67	92	7.2	12	51	95	13	61	91
B (VT)	$NRS_{G,\ -0.84SD}$	84	72	92	7.4	2	78	93	11	74	93
A/ISO	$NRS_{G,\ -0.84SD,\ x0.71}$	74	70	92	6.9	9	60	92	13	60	91

values only), and the alternative Method-A definitions with multiplicative derating. The Method-B analysis in Table 11 involves comparison of data from two different laboratories just as the Method-A analysis does: data from John Casali's lab at Virginia Tech in the former case and data from AFRL or European labs in the latter case. The four earplugs used in each analysis are the same or similar in materials and construction[20].

Table 11 shows that, when interlaboratory variability is included, Method B more closely achieves the protection rate target of 80% than derated Method A, though not by a large margin. The across-spectrum performance is also better for Method-B [lower $SD_N(P_{prot.})$], though the across-HPD performance is equivocal with a 12 to 14% higher protection rate, but a 2 to 3-dB worse underprotected equivalent level for the worst of the four earplugs. The contradictory findings are probably due to the tighter SD(A') value for the Method-A rating. Even though with the Method-A rating more wearers are over the fence of 85-dBA (as evidenced by the worse value for $P_{protected}$) they are not as far over the fence as with Method B, and hence a lower underprotected equivalent level is achieved for the worst case of the four earplugs. Overall, Method B's performance according to this limited analysis is slightly preferable to that of derated Method A, though not substantially better. Again, we state that this analysis, based on limited data, is tenuous and the aforementioned larger interlaboratory study would be valuable in order to clarify matters.

3.7 Waugh analysis using field data

Presuming our recommendation for a new rating based on Method-B data is adopted, what protection rates may be expected? We can estimate this with a Waugh analysis using field attenuation data for the Waugh population. Unfortunately, few such data exist and insufficient data are available to conduct the same evaluation on the Method-A rating proposed in Section 3.6. Table 12 shows the result of such a calculation using the only published field studies that include individual subject data needed to model the user population. The data used include only three HPDs, one premolded and two foam earplugs, all three of which are in the 20 HPD database. The field data are from NIOSH studies (Edwards et al., 1978 and 1982) and include a total of 84 subjects over three factories wearing the premolded earplug and 112 subjects over four factories wearing foam earplugs. Each subject's attenuation was measured on five

[20] The analysis was limited to only four earplugs because those were the only products sufficiently similar between the data sets. No comparable analysis of earmuffs was possible for lack of multiple-laboratory Method-B data on earmuffs that overlapped in model or in construction to the earmuffs in the 20-HPD database.

separate days; the average of those measurements is used to represent one member of the Waugh population.

The table shows that the protection rate evaluated using field data falls well shy of the target, with about half the population protected at *Q* = 80% and approximately two out of three protected at *Q* = 90%. This is not surprising if you recall that the goal of Method B as enunciated in the standard is to approximate "the upper limits of attenuation that can be expected for groups of occupational users." The Edwards et al. studies were conducted on "typical" programs that demonstrated performance falling short of those values.

Table 12 – Waugh analysis of recommended ratings computed from Method-B data using field data for three earplugs for the Waugh population. Noise spectra are the Ref300.

	Rating	$P_{prot.}$ %	P_{ideal} %	A'_{under} dBA
Q = 80%	NRS_A	48	42	96
	NRS_G	46	41	96
Q = 90%	NRS_A	64	53	95
	NRS_G	64	54	95
$NRS_{A,98}$		82	62	94
$NRS_{A,80}$ x 0.56		80	62	91

The underprotected equivalent level falls only 1 dBA in changing from the less conservative to the more conservative protection rate target. If a 2-SD adjustment is used in computing NRS_A then protection rates improve to 82% and the underprotected equivalent level decreases another 1 dBA. If ratings were reduced from their *Q* = 80% values by the preferred method of a multiplicative derating factor, in this case 0.56, the protection target is achieved and the underprotected equivalent level at 91 dBA is comparable to the value obtained using Method B to model the Waugh population (see Table 6). It is important to note, however, that if the analysis were based on a larger number of HPDs including some earmuffs, $P_{protected}$ would increase because of typically smaller differences between Method-B and field data for earmuffs than earplugs (Figure 4). It is also important to note that, while adjustments such as the 0.56-multiplier factor shown in the last row of the table can help ensure sufficient protection for the population and the Waugh analysis is a good way to test different adjustment possibilities, the above analysis includes field data on too few devices and HCPs (with their diverse degrees of HPD-use training) from which to recommend the amount of the adjustment. The only sure advice we can offer is that using a 2-SD or larger adjustment to the mean data in computing a rating is undesirable as it degrades the repeatability of rating values in repeated tests, as discussed in Section 3.2

4. RECOMMENDATIONS

In the forgoing analyses we have reviewed at length the types of data available to rate hearing protector performance, and the pros and cons of various rating systems. What is most clear is that definitive answers are elusive. The large degree of variability in the measurement processes (hearing-protector attenuation, use-time, and noise exposures) means that precise predictions of hearing damage risk for individuals are all but impossible, and even estimates for groups of users, are at best only indicative.

With respect to attenuation measurement process itself, we have shown that large differences in the measured data and the ratings computed therefrom arise as a function of the selected test method. For methods ranging from an optimized experimenter-fit approach using the currently mandated ANSI S3.19 standard to the subject-fit ANSI S12.6 Method-B procedure, the differences in rated values in decibels can be greater than 25 dB (see Footnote 5), and the differences in $P_{protected}$ can exceed 50% [see Table 5, $WC_H(P_{protected})$ for NRR vs. NRS_A].

Various arguments in favor of one approach or another can and have been constructed; the decision however is strongly influenced by one's perspective, philosophy, and assessment of the various factors. The argument for using Method-B data hinges on its better correlation with attenuation measured in the field (see Section 2.2) whereas the argument for Method-A is that it measures the inherent performance of the device obtainable when used by trained and experienced subjects. Proponents of each perspective see their respective viewpoints as the appropriate basis for a rating. In this report we have primarily worked with Method-B data because of its better correlation with field data, in particular because this allows it to properly rank order the relative performance of HPDs as measured on occupational users. We contend that if the purpose of the number on the package is to guide a buyer or a hearing conservationist in what can be realistically anticipated for a group of users, then subject-fit data like those derived from Method B are best suited to the job.

However, adoption of Method B poses some challenges. The values of ratings will decrease for almost all devices, more so for earplugs than for earmuffs (see Figure 8). In addition, even though Method B has been shown to be repeatable when viewed in terms of statistical confidence intervals, there is some evidence that the variability, measured as a percentage of the rated values, between laboratories and from test to test in a given laboratory might be greater than now exists (Berger, 2002). That raises problems for manufacturers whose emphasis is on consistent and fair ratings for products. We have considered that in our ratings selection (see Table 4 and the selection of Q = 80% and the NRS_A rating), and ANSI Working Group S12/WG11 will be addressing the issue in future revisions to the Method-B procedure. Additionally it is the reason that we include both Method-B and Method-A based recommendations.

To address the range of performance achieved by users who vary in diligence from casual to those who are highly trained and motivated, our proposal includes a dual rating system with both low- and high-values of protection reported. We anticipate that the lower value of the proposed ratings (Q = 80%) would be used to assess the adequacy of a particular HPD for a given noise environment in the absence of evidence to justify a higher value. It would be reasonable to allow a particular organization that employs personnel who might have received special or exemplary training because of exposure to extremely severe noise environments to conduct the Method-B laboratory measurements on a subset of the actual workers instead of using the normally required naïve test subjects. Then the test results and the computed 80/20 values would be directly representative for that group of users. Alternatively, if employers chose to use a fit-test type of system to directly check performance in the field (Michael, 1999), it would be appropriate to either allow use of those values, or to provide a means of interpolation between the 80/20 values to derive a rating more representative for that population. Similarly, systems that continuously monitor the noise dose of the individual worker by means of microphone(s) inside the HPD should be allowed as an alternative to the use of the labeled value (Burks and Michael, 2003).

With respect to the rating to be used to summarize and present the test data, the extensive analyses we conducted, utilizing over a dozen different metrics to assess the performance of the various ratings, found surprisingly few large differences between ratings. Clearly a class scheme, or ratings based on the current NRR yield values substantially inferior to the other methods (see Table 5), but once those were culled from the picture, the remaining single-number and multi-number ratings performed similarly for the vast majority of noises and hearing protectors; in extreme spectra or with unusual HPD attenuation spectra we did observe problems in the use of A – A' descriptors, hence the need for either a C – A' metric or an alternative graphical approach like the NRS_G. We decided against the C – A' approach, such as embodied in the ISO 4869-2 SNR value because of the inherent assumption about a particular shaped HPD attenuation spectrum that is intrinsic to that rating, as well as the potential confusion for users in being directed to subtract a single number from a C-weighted noise measurement to predict an A-weighted exposure. In the end our recommendation was motivated by our desire for a rating that:

- is easy to use (and hence worked with A-weighted sound levels),
- closely achieves the targeted protection rates,
- conveys to the user an indication of the range of performance that might be expected,
- most consistently achieves the protection rate in different noise spectra and for different HPDs,
- to the extent possible does not exacerbate, and conversely, potentially mitigates the impact of variability between test laboratories on the derived ratings.

4.1 The NRS_A and the rationale for two values on the primary label

The goals mentioned above lead to the selection of the $NRS_{A,80}$ and $NRS_{A,20}$ as the values to present either Method-B or Method-A data on the product's primary label, reflecting the attenuation to be anticipated for 80% and 20% of the situations (users X environments), respectively. The NRS_A is an A – A' descriptor that can be quickly and reliably subtracted from an A-weighted measurement. Averaging across noises and protectors it achieves the protection target within 1% (performing as well as the OBN), has a P_{ideal} within 2% of the OBN approach, works well across all HPDs, and only demonstrates shortcomings in the occasional spectra. Its $P_{Protected}$ in the noise in which it works most poorly drops just below 50% (by comparison the OBN remains slightly above 80%). To address this we propose the NRS_G (see below), whose performance as a rating closely tracks the "gold standard" OBN rating.

The selection of the $NRS_{A,Q}$ with the *Q* = 80% and 20% respectively serves a number of functions:

- It explicitly indicates that a range of performance is to be anticipated.
- It explicitly demonstrates products that offer less inter-subject variability since the range between the 80/20 numbers will be smaller.
- It diverts the attention of the buyer from a single value and the associated tendency to focus on the seeming "precision" of that value.
- It supports the rating of the product with a "safe" number that may appear low to some observers (the $NRS_{A,80}$ that 80% of the users can hope to achieve) while still indicating a much higher level of protection that is potentially attainable when everything is done just right.
- It draws attention to the possibility of overprotection.
- It may also encourage more careful fitting of hearing protection, especially among consumers who are buying products for their own use, by explicitly demonstrating what more careful application of the product can achieve.

In selecting the 80/20 pair other choices were considered such as 84/16, 90/10, or asymmetric pairs such as 90/50. The 84/16 approach has an appeal grounded in normal statistics; it represents plus or minus one SD and is in keeping with prior practice. We chose the 80/20 or 90/10 values primarily for clarity and simplicity for most users, since presenting results in terms of 4 out of 5 users, or 9 out of 10 users, is more understandable to many, especially those lacking a strong grasp of statistical concepts. Although somewhat more protective, the 90/10 approach was dismissed in favor of 80/20 because it creates a substantially larger range of values, and the $NRS_{A,90}$ approaches 0 dB for some devices, which would be confusing to many purchasers. Furthermore, as shown in Table 4, the SD of the rating value across sets of subjects is 15% larger, with Q = 90% compared to Q = 80%, suggesting that ratings resulting from repeat testing within and between laboratories would be more variable with the 90% statistic.

Selecting the 50% value for the upper end of the range ($NRS_{A,50}$) would assure that no high labeled values ranging from the mid to upper 30s would be reported, as would otherwise be the case with $NRS_{A,20}$. However use of $NRS_{A,50}$ would create an asymmetric pair of ratings that again might be confusing to some. Another reason for consideration of the 50% value is that some users may contend that they have excellent programs and thus should be allowed to use the upper of the two values on the label (be it $NRS_{A,50}$ or $NRS_{A,20}$ as representative for their groups), and with the $NRS_{A,50}$ there would be less risk in this approach. However, that is not the intended application of the upper number in the range and we cannot protect against all manner of misuse. An advantage of the $NRS_{A,20}$ is that it is easier to conceptualize in terms of the type of user it represents (see discussion in Sections 4.3 and 4.4 on the recommended label).

One of the virtues of the 80/20 range is that the current NRRs will fall in most cases between the 80/20 numbers so that there will be no dramatic change in the labeled values. We hope this will be easier to understand and accept. The relationship between the new numbers and the old numbers for the 20 representative HPDs that we studied is presented in Figures 8 and 14; overall (averaging together earplugs and earmuffs) the $NRS_{A,80}$ computed from Method-B data will be about 5-dB less than labeled NRRs. However, when one considers that the labeled NRRs are meant for use with C-weighted values and must be reduced from 2.5 to 7 dB for use with A-weighted sound levels (see Sections 2.2 and 3.1) the $NRS_{A,80}$ and the NRR as used today are about the same averaged across all products. For earplugs though, the differences will be larger and the $NRS_{A,80}$ values could be as much as 10-dB less for some products. On the other hand the $NRS_{A,20}$ values will substantially exceed the NRRs for earmuffs, and on average exceed them by a few decibels for earplugs as well.

Similar protection rates can be achieved by using Method-A data with the same 80/20 range and the inclusion of a multiplicative correction factor of 0.71 for Q = 80% and 0.88 for Q = 20%. The multiplicative correction is required to reduce the rating values that result from applying the same SD adjustment as we recommend for Method-B data to Method-A data with its higher values of mean attenuation and lower SDs. The alternative (using a 2-SD adjustment for Method A) offers no improvement in protection rates or the other metrics while reducing the repeatability of a Method-A rating to no better than can be computed from Method B (see Section 3.6). The downside of the Method-A based approach is that the average exposures for those worst-case protected users will be 3 - 4 dB higher than with a Method-B rating, and as discussed in Section 3.6 with respect to Figure 14, the Method-A ratings transpose the relative overall average earplug and earmuff attenuation values from what has been observed in field studies.

We considered the possibility of providing different corrections for earplugs and earmuffs, or even a more detailed type-specific correction like NIOSH (1998) currently recommends but contend that without testing each product it is not possible to fairly or appropriately devise such detailed deratings.

4.2. An alternative "high-precision" value for the secondary label

Though for the majority of noises and HPDs the precision of the NRS_A is sufficient for the data with which it is utilized, when C – A values exceed approximately 6 dB (~10% of the Ref300, though about 60% of the AF 50 + CA 20) protection rates can fall below 70% and underprotected exposure levels increase by about 1.5 dB beyond what would be experienced with an OB estimation (see Figures 11 and 12, and Table 7). For such noises, those who wish to make a more accurate prediction should be encouraged to turn to the information on the secondary label presented in terms of the NRS_G values. An example is provided in Section 4.5.

One could, of course, choose the "gold standard" OB approach, but the principal problem therein is that the instrumentation needed to obtain long-term OB dosimetry data is neither commonly available nor easily used. Single-point or short-duration OB sound-level measures are relatively easy to obtain, but most hearing conservationists would argue that time-weighted average exposures or doses are more representative and reliable to apply (Royster, Berger, and Royster, 2000). Thus for most applications one is better off using A- or C-weighted measures and hence a single number to subtract from such values is needed. The NRS_G provides such a number, but as explained in this report, derived in such a manner as to utilize the spectral attenuation values of the hearing protector more fully, and to closely approximate the results obtained with the OB approach. In this regard, the NRS_G performs like its forbearers the HML (Lundin, 1986) and the USAF 5-number approach (McKinley, 2001 based on Johnson and Nixon, 1974), with the principal differences being its simple graphical format, and the fact that it is computed from the individual subject mean-attenuation values instead of from the mean and SD of the attenuation across all subjects. This more closely models the effects of the individual variability in each noise spectrum.

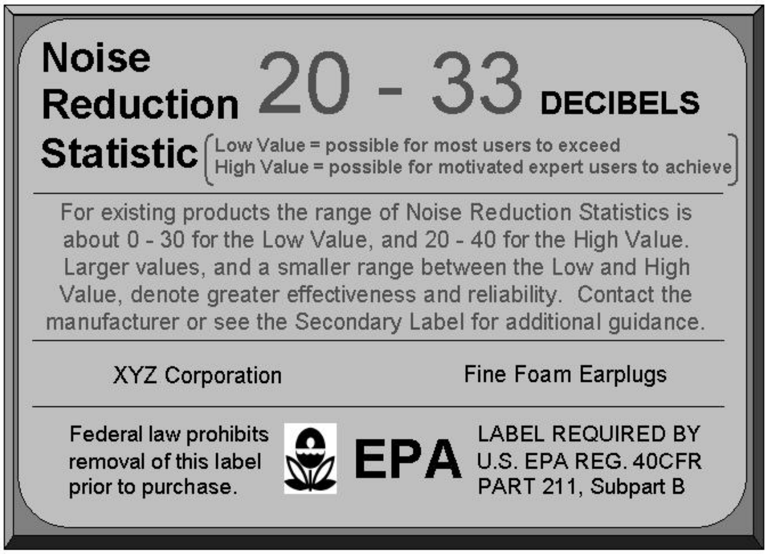
4.3. Proposed format for the primary label

The format for the proposed primary label is shown in Figure 15 with the newly proposed range and accompanying explanatory wording shown in gray. The attenuation rating is labeled the Noise Reduction Statistic (NRS) to clearly distinguish it from the prior mandated NRR. We believe that within the intent of the law, a value designated the NRS can be considered a "noise reduction rating" though it is not explicitly designated as the "Noise Reduction Rating." Use of a new term would create less confusion during the changeover from an old to a new labeling requirement. However, the fact that the proposed label includes two numbers instead of one should also make it apparent to even the casual observer that something has changed.

Note that in parentheses below the NRS values an explicit statement of their meaning is provided. The low value, 20 dB in this example, is a value that should be "possible for most users to exceed." The intent is that the user can with some degree of confidence expect to obtain such protection. The word "possible" indicates that it is not guaranteed. The high value, 33 dB in this example, is only "possible for motivated expert users to achieve." The intent here is to convey the understanding that the user must exert great care and diligence, and be skilled in the use of the product to hope to obtain such a level of protection.

In the next section of the label an indication of the range of values that are expected for existing products is provided. This is based on the data in this report and may require modification once a larger database is available. However, the approximate nature of the range suggests that these estimates may be sufficient. The user is also told the larger ratings indicate more protection, and a smaller range between the numbers is desirable as well. Finally, the user is directed to the secondary label for additional guidance, and it is here that they will find the NRS_G and other information.

Figure 15 – Proposed primary label illustrating $NRS_{A,20}$ and $NRS_{A,80}$ values based upon Method-B data, and brief explanation of their meaning.



Alternatively, should Method-A data be selected, the values on the label would still be $NRS_{A,80}$ and $NRS_{A,20}$, though their computation would include the multiplicative constant referred to earlier. The low- and high-value descriptions would remain the same as for the $NRS_{A,80}$/$NRS_{A,20}$, as would the paragraph of text on the label itself.

4.4. Proposed material for the secondary label

We recommend that the secondary label, as in the current regulation, be required to provide substantial additional information to guide the user. Because of space limitations on product packaging and because of the additional information that we are suggesting beyond the secondary label information that is now required under the law, the secondary label may need to be separated into a secondary and tertiary label, part of which would appear on the product packaging or dispenser box and the remainder of which would be available on the web or by mail. First and foremost the secondary label must provide guidance in application of the NRS. We propose:

Select either the Low or High Value as representative for your use. Subtract that value from either an A-weighted sound level (dBA), or a time-weighted average noise exposure in dBA as follows:

1. *The noise level or noise exposure is 92 dBA.*
2. *The NRS (Low Value) is 20 dB.*
3. *Most users should be protected to a level of 72 dBA.*

Tip: For greater precision in estimating protection, especially in noises with substantial low-frequency energy, contact the manufacturer for the Noise Reduction Graph for this product.

At a minimum, the secondary label also should contain the following information, with the first four items appearing on the secondary label and the remainder perhaps being provided on a tertiary label:

- Complete instructions on how to properly fit and care for the HPD and the importance of following the instructions in order to achieve the performance the device is capable of providing.
- A warning that “the hearing protector must be worn at all times in noisy surroundings for proper protection to be achieved.”
- Additional discussion of the meaning of the two values on the label (see *Using the range of values*, below).
- A graph depicting the $NRS_{G,80}$ and $NRS_{G,20}$ values (see Figure 6, and proposed wording in *Applying the NRS Graph*).
- A table containing the octave-band means and SDs at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz.
- Discussion of overprotection (see *Overprotection*, below).
- The importance of additional selection criteria besides attenuation per the recommendations of the National Hearing Conservation Association (NHCA) *Task Force on Hearing Protector Effectiveness* (Royster, 1995; and see *Hearing protector selection criteria*, below).

Following is the suggested wording for the third, fourth, sixth, and seventh bulleted items listed above.

Using the range of values:

The noise ratings on the label are based on laboratory tests that have been shown to simulate actual use. However, it is not possible to reliably predict the protection achieved by a selected group of users or a given individual. The range of values provides an indication (using the low number) of the protection that can be achieved in most hearing conservation programs by about 4 out of 5 typical users, and (using the high number) an indication of what a well trained and motivated individual user can obtain. The range of values makes clear the importance of the behavior of the user in achieving the maximum protection the device can provide.

Applying the NRS Graph:

For noises that have a rumbling, thunderous, or heavy sound to them, for sounds that are from air-moving equipment, for passengers in moving vehicles, or if the measured difference between A- and C-weighted sound levels exceeds 6 dB, use of this procedure is recommended. Find the C – A value of the noise on the bottom axis, read up to the appropriate line for the low or high protection rating, and then over to the left axis to find the protection that is to be subtracted from the A-weighted sound level.

Overprotection:

When selecting a hearing protector, more protection is <u>not</u> always the best choice. In moderate noise environments, or when there is a critical need to hear communication and warning signals, and for individuals who already have a moderate high-tone hearing loss, a less protective device that still provides adequate noise reduction may be desirable. If subtraction of the high number on the label from the noise measurement leads to estimated levels well below 70 dBA, this suggests that overprotection may occur.

Hearing protector selection criteria:

The most critical consideration in selecting and dispensing a hearing protector is the ability of the wearer to achieve a <u>comfortable</u> noise-blocking seal that can be consistently maintained during all noise exposures. Additional important issues include: the device’s noise reduction, the person’s noise exposure, user preferences and communication needs, hearing ability, compatibility with *other safety equipment such as eyewear and respirators, physical disabilities that make use of devices difficult, and climate and other working conditions.*

Differences between hearing protector ratings for two devices, of less than 3 dB are not important.

4.5. How to apply the ratings

In this section we provide additional guidance concerning the ratings. This information is not intended for the label, but could be published in future documents depending upon the course of EPA's deliberations, in order to offer guidance and explanatory materials to professional hearing conservationists.

As explicitly stated in the example on how to use the NRS (Section 4.4), the decibel value of that number can be subtracted from either a sound-level measurement or a TWA. The choice depends on how the user characterizes the noise hazard of the environment. Likewise, although we have not addressed it explicitly, the NRS can be subtracted from impulse-noise measurements. Data suggest that, if anything, attenuation values based on steady-state measurements will generally underestimate the actual protection achieved in impulsive environments, even more so for the case of intentionally level-dependent HPDs that are designed to provide increasing attenuation with sound level (Berger, 2000, p. 426-427).

As we have discussed, the NRS is intended for application to A-weighted measurements. This choice was made because of the evidence that the existing NRR, intended for subtraction from C-weighted levels, is often misused. In fact, the current mandated EPA wording on the packaging (EPA, 1979) encourages misuse since, although a cautionary note about use of C-weighted levels in low-frequency environments is included, the worked example that is required illustrates use of the NRR with an A-weighted sound level. The need for an A-weighted friendly rating is obvious, especially in the consumer market where the buyer is unlikely to know A, C, or any other weighting, let alone the meaning of a decibel. Guidance then must be provided on when to use the more accurate NRS_G. That is found in the tip, listed below the worked example in Section 4.4 where the user is directed to the graphical approach. To use this method the user must measure both the A- and C-weighted levels, or exposures, to compute the C – A value, and then using the chart read off the appropriate protection that can be subtracted from the A-weighted measurement. As with the primary label, two values are provided, this time in the form of two lines on the chart. Again they represent the 80% and 20% protection rates.

In Section 2.4 we stated that one of our goals for a new rating is to have it be comprised of two numbers that convey the range of protection an HPD provides. Our motivation for this is to communicate the level of precision in device performance and the measurement thereof, so that people do not base purchase decisions largely on insignificant differences in noise reduction performance, but also consider important factors such as user preference, hopefully leading to greater worker acceptance of using HPDs. Providing two values that define the range of reduction offers another advantage though: it can provide the professional hearing conservationist information useful to refine the choice among HPDs that suitably balance over- and underprotection.

4.6. The Method-B / Method-A conundrum and the need for an interlaboratory study

Our analyses have led to the conclusion regarding the need for a dual number A – A' rating and a secondary graphical rating that utilizes additional spectral information for a more accurate prediction. However, our recommendations are not as clear-cut with respect to the type of data to utilize as the basis for the computation of the new ratings. The analyses emphasized the use of Method-B data because as shown in Section 2.2 those values provide more useful estimators of field performance. Nevertheless concerns have surfaced regarding the reproducibility of Method-B results (Berger, 2002) and hence a Method-A approach was also evaluated. We found that with suitable corrections similar predictions could be developed using the Method-A results, albeit with less accuracy for certain noise spectra and HPDs.

Our conclusions are tempered by the lack of a sufficient body of data on which to conduct our analyses. Few Method-B data are available in North America except for the results from the laboratory managed by the second author. With respect to Method-A testing almost no data are available at this time; we had to supplement the Method-A data we had available for analyses by incorporating European results from various laboratories, tested to the similar ISO 4869-1 test standard. More solid recommendations could be provided were there to be available a homogeneous set of data comparing the two standards in a group of laboratories. Thus we recommend that an interlaboratory study be conducted under the auspices of S12/WG11, with funds and support from EPA and NIOSH.

The study should include at a minimum five laboratories to be selected from the facilities that are currently accredited under NVLAP (E•A•RCAL^{SM} Laboratory of Aearo Company, the laboratory of Bacou-Dalloz, and Michael & Associates, Inc.), and include as many of the government (NIOSH, Wright-Patterson AFB, USAARL Alabama) and other independent facilities (Virginia Tech) as possible. The study should include both Methods A and B[21], a selection of preferably 10 HPDs, but at least eight (to include 1 roll-down foam plug, 1 premolded plug, 1 custom molded plug, 1 semi-insert, and 1 small-, medium-, and large-volume earmuffs with at least one in the behind-the-head wearing position). Many of the facilities should be able to self-fund their participation so that only one or two would require monies to participate, probably on the order of $50,000 per laboratory.

The conduct of such a study should not hold up the EPA rulemaking process, since the testing could be conducted subsequent to the proposed rulemaking and the results made available prior to the final promulgation. Planning and conduct of such a study should be able to occur within about 18 months.

Regardless of the decision concerning additional testing, we strongly recommend that either a Method-B or Method-A based rating be selected, <u>not</u> both. The requirement for testing to both procedures would be unduly burdensome on manufacturers, not only in the initial roll out of the new ratings, but especially during the development cycle of new products. It would also be needlessly confusing to the consumer. Furthermore, EPA must make the decision; a manufacturer should not have the choice concerning the type of data to use to rate their products.

5. CLOSING REMARKS

The foregoing analyses and discussion have extensively examined the rating of hearing protection devices. HPDs are an important means, perhaps the principal means available for hearing conservationists to protect the hearing of their noise-exposed employees. In non-occupational settings hearing protection use is proliferating as well. Specifiers and buyers of such products need to know how much protection they can anticipate. Providing an answer is not easy. What is clear, however, is that the current values on product packaging, though they have raised awareness of the need for hearing protector attenuation, have done little to provide valid guidance in selection of products. Our hope is that the ratings and methods we have proposed, together with the findings of a needed interlaboratory study,

---

[21] Since the publication of ANSI S12.6-1997 and its reaffirmation, various refinements to the procedures, especially Method B have been suggested and some or all of those should be incorporated in modified Method-A and B procedures to be used for an interlaboratory comparison.

can be incorporated in an improved hearing protector labeling regulation that will guide users in a more balanced and appropriate selection of products, and in turn lead to a reduction of the incidence of noise-induced hearing loss.

6. REFERENCES

1. ANSI (1957). "Method for the Measurement of Real-Ear Attenuation of Ear Protectors at Threshold," American National Standards Institute, Z24.22-1957 (R1971), New York, NY.
2. ANSI (1974). "Method for the Measurement of Real-Ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs," American National Standards Institute, S3.19-1974 (ASA STD 1-1975), New York, NY.
3. ANSI (1984). "Method for the Measurement of the Real-Ear Attenuation of Hearing Protectors," American National Standards Institute, S12.6-1984, New York, NY.
4. ANSI (1997). "Methods for Measuring the Real-Ear Attenuation of Hearing Protectors," American National Standards Institute, S12.6-1997, New York, NY.
5. Backshall, D. (2000). Personal communication of previously unpublished data.
6. Berger, E. H. (1989). "Exploring Procedures for Field Testing the Fit of Earplugs," in *Proceedings, 1989 Industrial Hearing Conservation Conference*, Off. Eng. Serv., Univ. Kentucky, Lexington, KY, 7-10.
7. Berger, E. H. (2000). "Hearing Protection Devices," in *The Noise Manual, 5th Edition*, edited by E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll, and M. Layne, Am. Ind. Hyg. Assoc., Fairfax, VA, 379-454.
8. Berger, E. H. (2002). “Report #43 – Information for review for upcoming meeting of WG11,”
9. Berger, E. H., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (1998). "Development of a New Standard Laboratory Protocol for Estimating the Field Attenuation of Hearing Protection Devices. Part III. The Validity of Using Subject-Fit Data," J. Acoust. Soc. Am. 103(2), 665-672.
10. 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. Axlesson, H. Borchgrevink, R. P. Hamernik, P. Hellstrom, D. Henderson, and R. J. Salvi, Thieme Medical Pub., Inc., New York, NY, 361-377.
11. Berger, E. H. and Kieper, R. W. (2000). "Ongoing Experience with Laboratory-Based Subject-Fit REAT Methods for Measuring Hearing Protector Attenuation," J. Acoust. Soc. Am. 108(5), Pt. 2, p. 2620.
12. Berger, E. H. and Royster, L. H. (1996). "In Search of Meaningful Measures of Hearing Protector Effectiveness," Spectrum Suppl. 1, 13, p. 29.
13. Botsford, J. H. (1973). "How to Estimate dBA Reduction of Ear Protectors," Sound and Vibration 7(11), 32-33.
14. Brinkmann, K. and Serra, M. R. (1982). "Acoustical and Mechanical Properties of Earmuff Type Hearing Protectors and Their Testing," in *Personal Hearing Protection in Industry*, edited by P. W. Alberti, Raven Press, New York, NY, 133-150.
15. Burks, T. A. and Michael, K. L. (2003). “A New Best Practice for Hearing Conservation: The Exposure Smart Protector (ESP), in *Proceedings of Noise-Con 2003*, edited by D. K. Holger and G. C. Maling, Jr., Inst. Noise Control Eng. USA, Washington, DC, paper 009.
16. CEN/TC (2003). “Hearing Protectors – Recommendations for Selection, Use, Care and Maintenance – Guidance Document," European Committee for Standardization, CEN/TC 159 prEN 458, Brussels.
17. CSA (2002). "Hearing Protection Devices - Performance, Selection, Care, and Use," Canadian Standards Assoc., Z94.2-02, Rexdale, Ontario.
18. Earshen, J. J. (2000). "Sound Measurement: Instrumentation and Noise Descriptors," in *The Noise Manual, 5th Edition*, edited by E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll, and M. Layne, Am. Ind. Hyg. Assoc., Fairfax, VA, 41-100.

19. Edwards, R. G., Hauser, W. P., Moiseev, N. A., Broderson, A. B., and Green, W. W. (1978). "A Field Investigation of Noise Reduction Afforded by Insert-Type Hearing Protectors," National Institute for Occupational Safety and Health, U.S. Dept. of HEW, Rept. No. 79-115, Cincinnati, OH.
20. Edwards, R. G., Broderson, A. B., Green, W. W., and Lempert, B. L. (1982). "A Second Field Investigation of Noise Reduction Afforded by Insert-Type Hearing Protectors," National Institute for Occupational Safety and Health, U.S. Dept. of HHS, Rept. No. 210-81-3001, Cincinnati, OH.
21. Else, D. (1973). "A Note on the Protection Afforded by Hearing Protectors - Implications of the Energy Principle," Ann. Occup. Hyg. 16, 81-83.
22. EPA (1979). "Noise Labeling Requirements for Hearing Protectors," Environmental Protection Agency, Fed. Regist. 44(190), 40CFR Part 211, 56130-56147.
23. Gauger, D. (2003). “Testing and Rating of ANR Headsets,” U. S. EPA Workshop on Hearing Protector Devices, Washington, DC.
24. Hall, J. A. (2003). “Hearing Protection Metrics: U. S. Air Force Perspectives,” U. S. EPA Workshop on Hearing Protector Devices, Washington, DC.
25. Hall, J. A. (2004). Personal communication.
26. ISO (1990). "Acoustics - Hearing Protectors - Part 1: Subjective Method for the Measurement of Sound Attenuation," International Organization for Standardization, ISO 4869-1:1990(E), Switzerland.
27. ISO (1994). "Acoustics - Hearing Protectors - Part 2: Estimation of Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn," International Organization for Standardization, ISO 4869-2:1994(E), Switzerland.
28. Johnson, D. L. and Nixon, C. W. (1974). "Simplified Methods for Estimating Hearing Protector Performance," Sound and Vibration 8(6), 20-27.
29. Karplus, H. B. and Bonvallet, G. L. (1953). "A Noise Survey of Manufacturing Industries," Ind. Hyg. Qtrly., December, 235-263.
30. Kroes, P., Fleming, R., and Lempert, B. (1975). "List of Personal Hearing Protectors and Attenuation Data," National Institute for Occupational Safety and Health, U.S. Dept. of HEW, Rept. No. 76-120, Cincinnati, OH.
31. Lundin, R. (1986). "New Nordic Draft Standard for Calculation of Three Attenuation Parameters for a Hearing Protector and How to Use Them in Practice," in *Proceedings of Inter-Noise 86*, edited by R. Lotz, Noise Control Foundation, Poughkeepsie, NY, 553-558.
32. McKinley, R. (2001). Personal communication.
33. McQueen, G. H., Cameron, A. J., King, R. B., Mannam, J., Reilly, R. N., Stafford, R. G., and Wilson, K. J. (1969). "Good Health," Occup. Health Branch, Dept. of Public Health, Adelaide, Australia.
34. Michael, K. (1999). “Measurement of Insert-Type Hearing Protector Attenuation on the End-User: A Practical Alternative to Relying on the NRR,” Spectrum 16(4), 13-17.
35. Murphy, W. J. (2003). “Deriving a new NRR from ANSI S12.6B method, interlaboratory reproducibility of data, precision of the data,” U. S. EPA Workshop on Hearing Protector Devices, Washington, DC.
36. Murphy, W. J., Franks, J. R., Berger, E. H., Behar, A., Casali, J. G., Dixon-Ernst, C., Krieg, E. F., Mozo, B. T., Royster, J. D., Royster, L. H., Simon, S. D., Stephenson, C. (2004). "Development of a New Standard Laboratory Protocol for Estimation of the Field Attenuation of Hearing Protection Devices: Sample Size Necessary to Provide Acceptable Reproducibility." J. Acoust. Soc. Am. 115, 311-313.
37. NIOSH (1998). "Criteria for a Recommended Standard - Occupational Noise Exposure, Revised Criteria," Natl. Inst. for Occup. Safety and Health, DHHS (NIOSH) Pub. No. 98-126, Cincinnati, OH.
38. OSHA (1983). "Occupational Noise Exposure; Hearing Conservation Amendment; Final Rule," Occupational Safety and Health Administration, 29CFR1910.95 Fed. Regist. 48(46), 9738-9785.
39. OSHA (1990). "OSHA Instruction CPL 2-2.20B, February 5," in *OSHA Technical Manual*, Government Institutes, Inc., Rockville, MD, p. 4-1 - 4-15.

40. Royster, J. D., Berger, E. H., Merry, C. J., Nixon, C. W., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Mozo, B. T., Ohlin, D., and Royster, L. H. (1996). "Development of a New Standard Laboratory Protocol for Estimating the Field Attenuation of Hearing Protection Devices. Part I. Research of Working Group 11, Accredited Standards Committee S12, Noise," J. Acoust. Soc. Am. 99(3), 1506-1526.

41. Royster, L. H. (1995). "In Search of a Meaningful Measure of Hearing Protector Effectiveness: Recommendations of the NHCA's Task Force on Hearing Protector Effectiveness," Spectrum 12(2), p. 1 and 6-13.

42. Royster, L. H., Berger, E. H., and Royster, J. D. (2000). "Noise Surveys and Data Analysis," in *The Noise Manual, 5th Edition*, edited by E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll, and M. Layne, Am. Ind. Hyg. Assoc., Fairfax, VA, 165-244.

43. SAS (1995). *SAS/QC Software: Usage and Reference, Version 6, First Edition, Volume 1*, SAS Institute Inc., Cary, NC, p. 847.

44. SA/SNZ (1999). "Acoustics - Hearing Protectors," Standards Australia and Standards New Zealand, 1270:1999, Strathfield, NSW.

45. Sutton, G. J. and Robinson, D. W. (1981). "An Appraisal of Methods for Estimating Effectiveness of Hearing Protectors," J. Sound Vib. 77(1), 79-81.

46. The MathWorks (2000). *Matlab® Technical Computing Language Software (Release 13)*, The MathWorks Inc., Natick, MA.

47. Thomas, W. C. and Casali, J. G. (1995). "Instructional Requirements for Using the HML and NRR Methods for Estimating Protected Exposure Levels under Hearing Protectors," Auditory Systems Laboratory, Virginia Tech, Rept. No. 3/1/95-1-HP; ISE Rept. 9502, Blacksburg, VA.

48. Ward, W. D. (1982). "Risks to Hearing: The Effect of Intermittent Exposure," in *Personal Hearing Protection in Industry*, edited by P. W. Alberti, Raven Press, New York, NY, 35-49.

49. Waugh, R. (1973). "dBA Attenuation of Ear Protectors," J. Acoust. Soc. Am. 53(2), 440-447.

50. Waugh, R. (1976a). "Calculated In-Ear A-Weighted Sound Levels Resulting From Two Methods of Hearing Protector Selection," Ann. Occup. Hyg. 19, 193-202.

51. Waugh, R. (1976b). "Investigation of Sound Level Conversion as a Means of Rating Ear Protector Performance," Am. Ind. Hyg. Assoc. J. 37(4), 239-245.

52. Waugh, R. (1984). "Simplified Hearing Protector Ratings - An International Comparison," J. Sound Vib. 93(2), 289-305.

ANNEX A – METHOD-B, ISO 4869, and S3.19 DATA FOR THE HPDs in this Study

In this Annex we include various graphs describing the attenuation data for the HPDs that were used in the foregoing evaluation. All of these data, as well as the additional data used in the analyses in Section 3.6 are available in electronic format from the authors.

Figures A1 – A4 provide the mean Method-B real-ear attenuation data for a group of 47 HPDs that were considered for inclusion in this study. The bolded lines represent the 20 devices selected according to the criteria described in Section 2.1.

Figure A5 depicts histograms for the 20 selected products to demonstrate the types of distributions that are observed in Method-B testing. Per the standard, earplugs are tested using 20 subjects and earmuffs using 10.

Figure A6 presents the APVs for the 20 HPDs, computed from the both the Method-B and ISO 4869-1 data, with a 1-SD correction, and from the S3.19 data with the required 2-SD correction.

Figure A7 presents the NRS_G at the 20^{th} and 80^{th} for the 20 HPDs. This illustrates the graphs that would appear on the secondary labels if the recommendations of this report are followed.

Figure A1 - Foam and premolded earplugs: included products in bold; dashed line is flat plug.

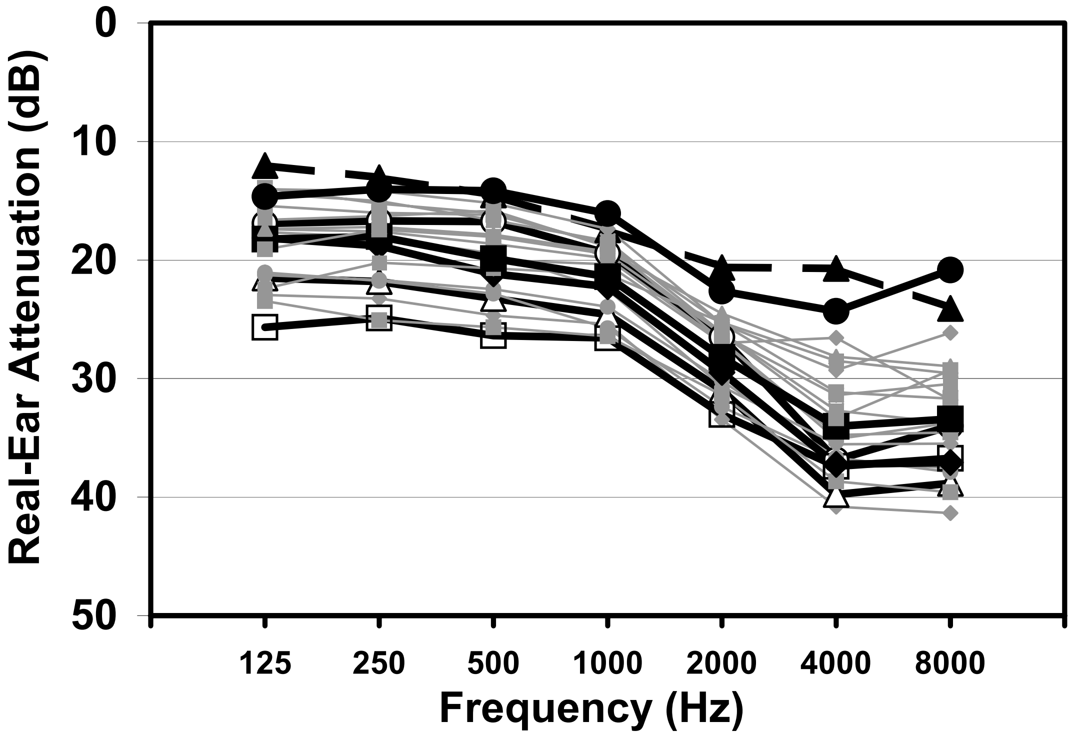


Figure A2 - Semiinserts: included products in bold.

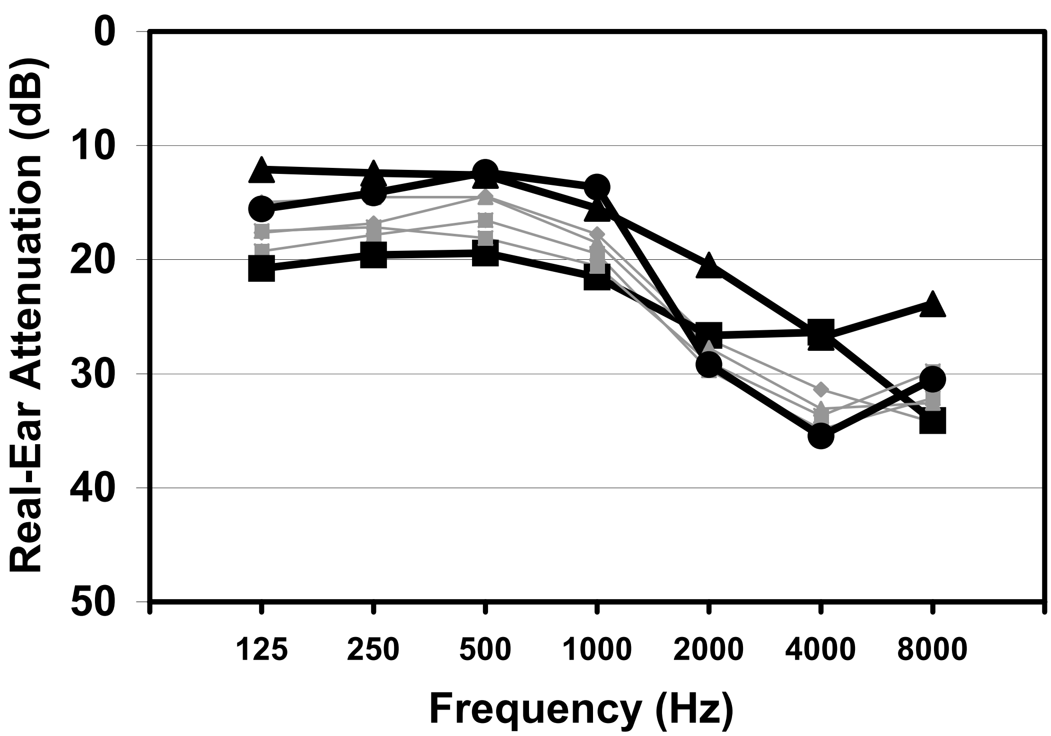


Figure A3 - Earmuffs: included products in bold; dashed line is flat attenuator.

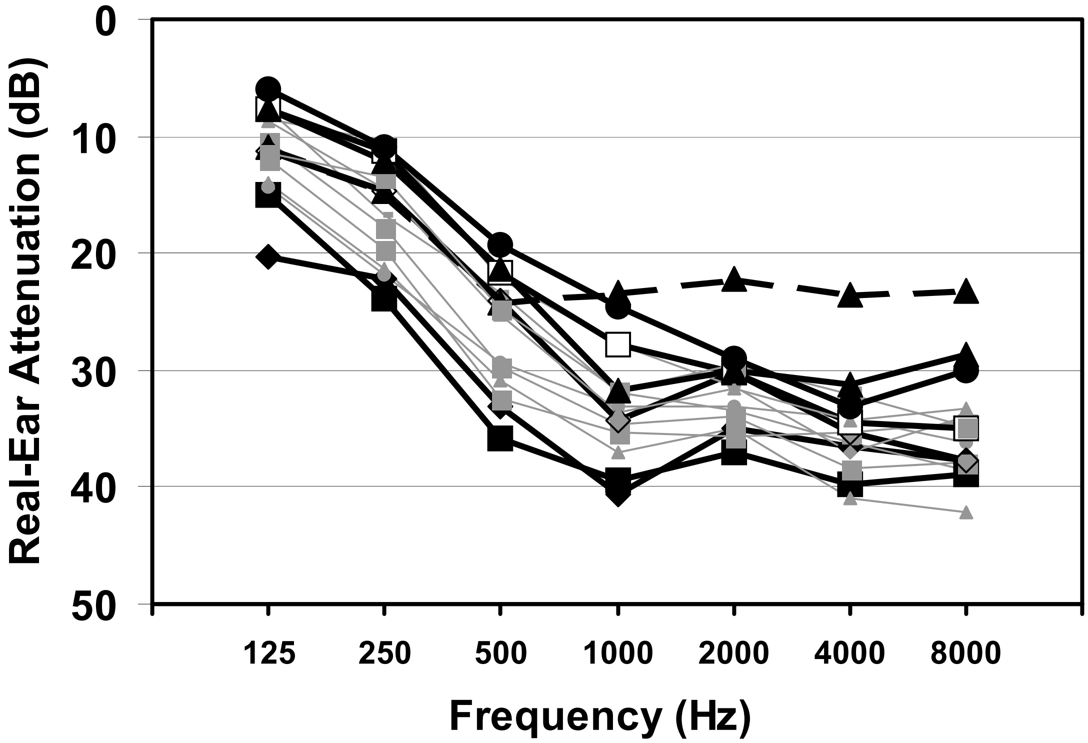


Figure A4 - Cap-attached earmuffs and dual-protection: included products; dashed line is dual protection.

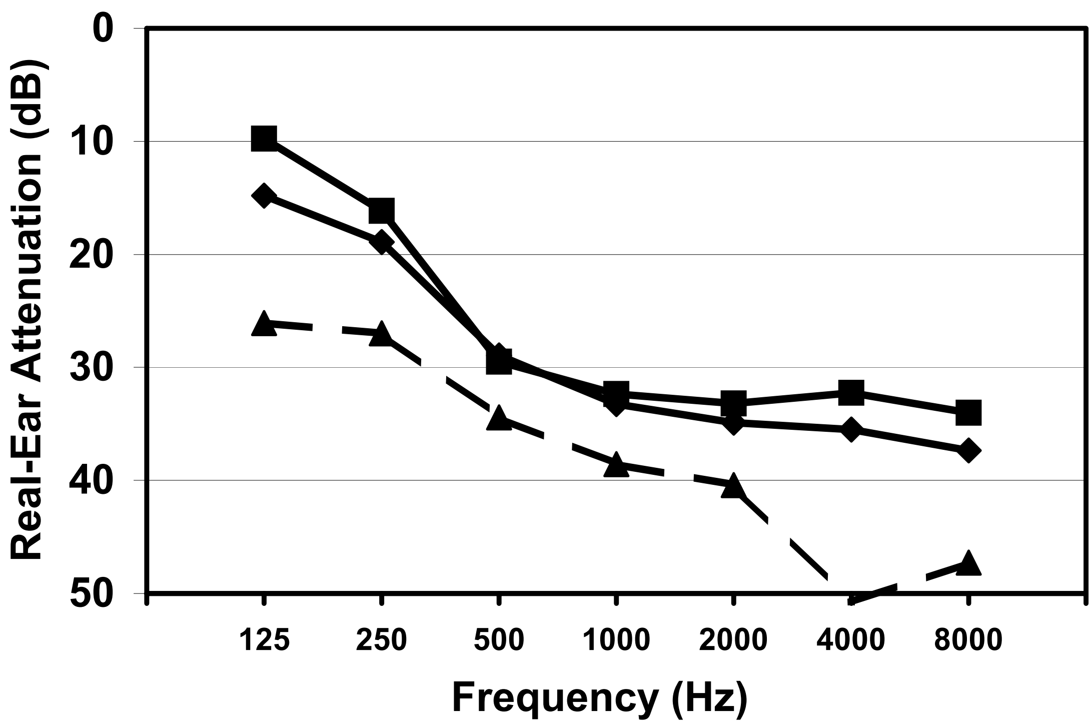


Figure A5 – Histogram distributions for the Method-B test results for the 20 HPDs showing distributions with various degrees of normality and bimodality. The quantity plotted on the X-axis is the protection in pink noise for each subject. The overlaid line is the normal distribution fit to the data.

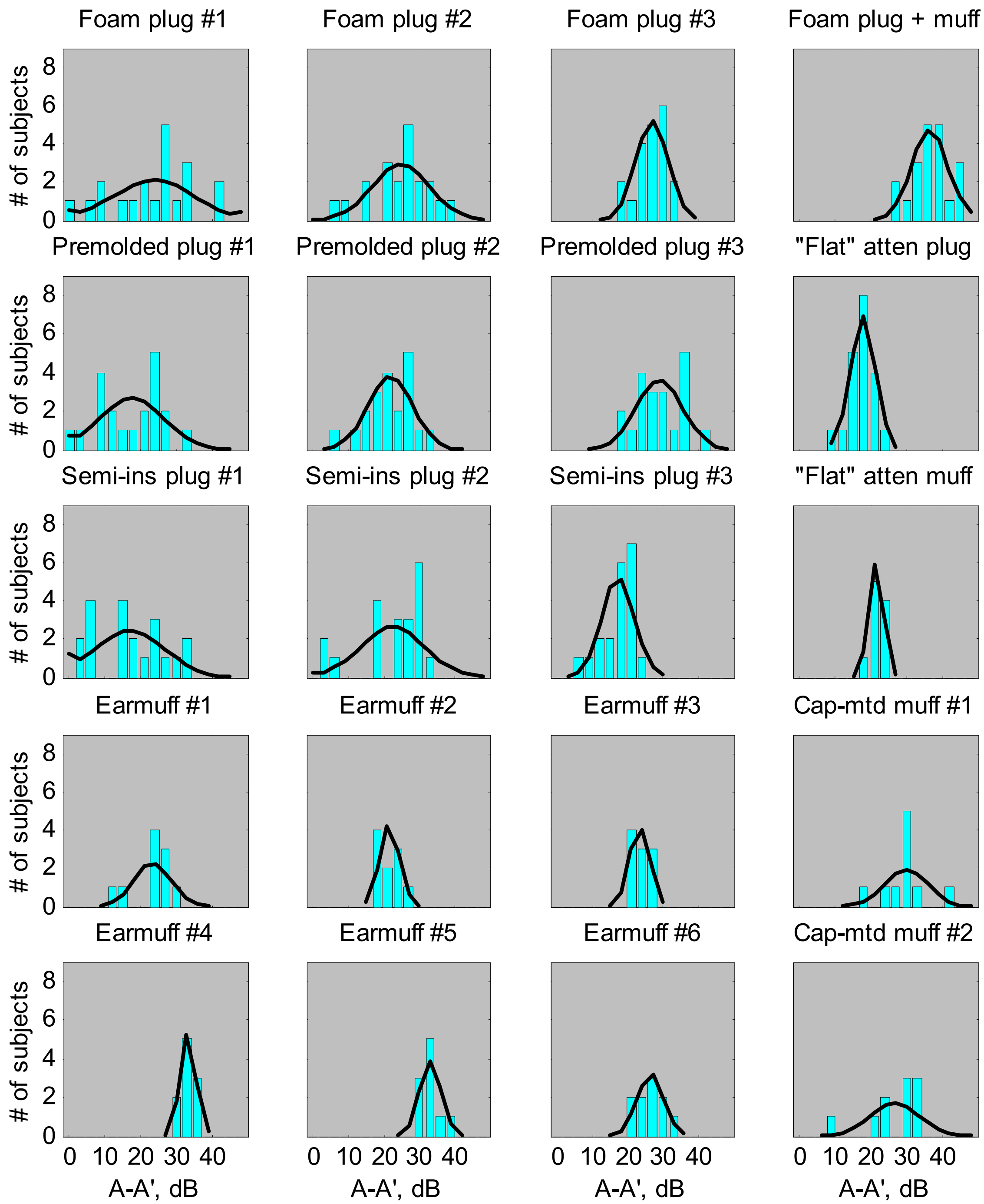


Figure A6 – Assumed protection values for the 20 HPDs for Method B, ISO 4869, and ANSI S3.19, computed with 1 SD, 1 SD, and 2 SDs respectively.

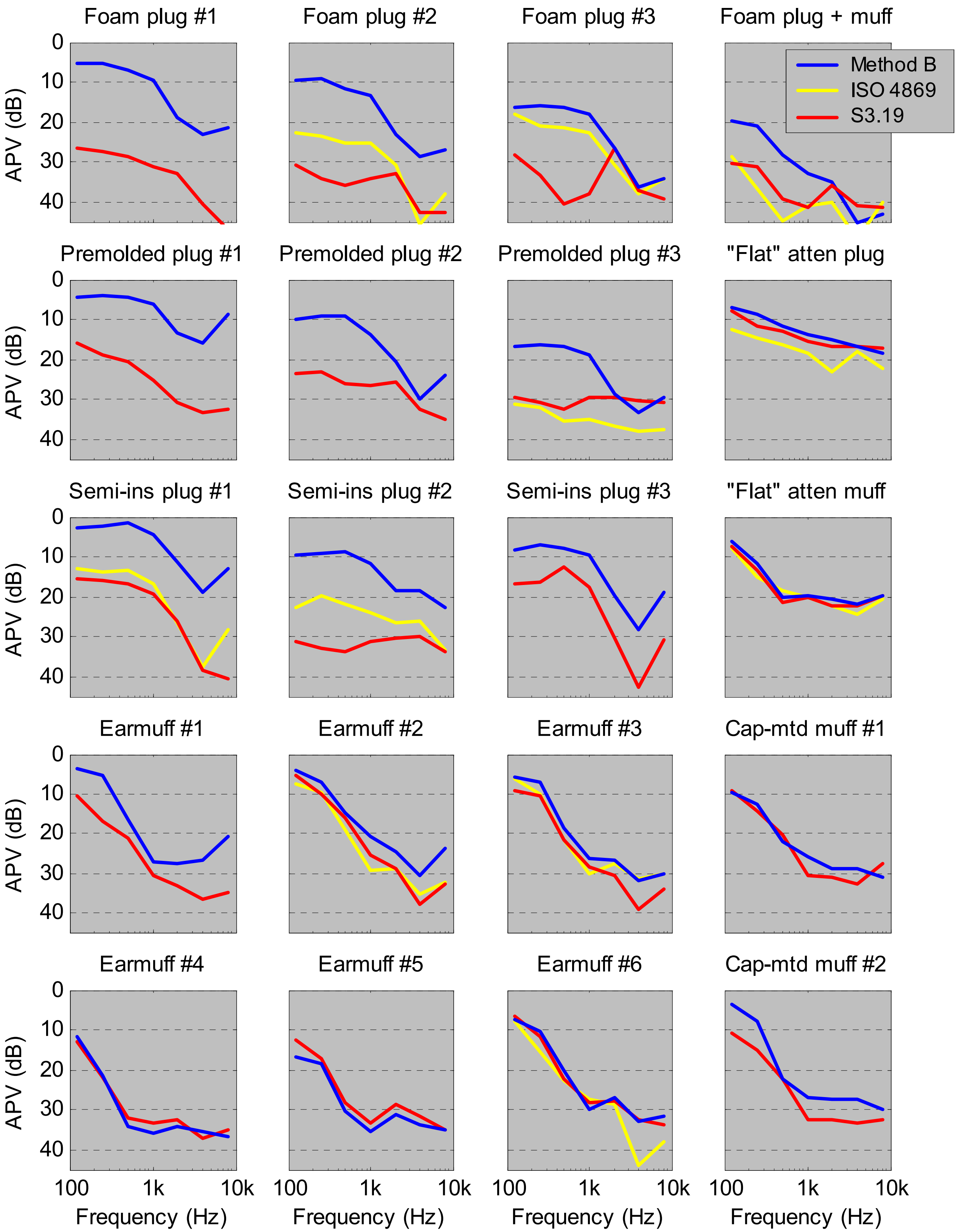
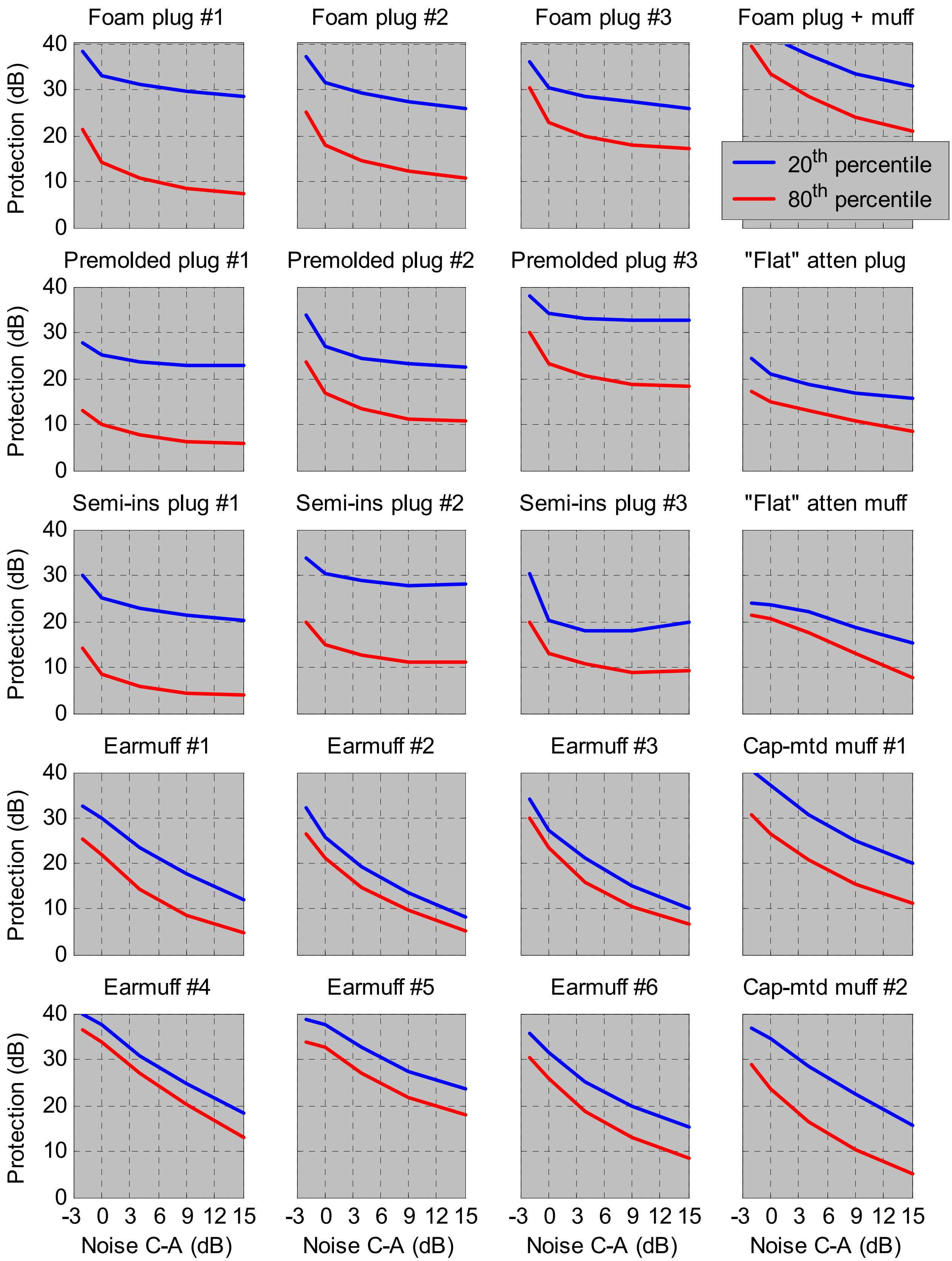


Figure A7 – Noise Reduction Statistic, graphical (NRS_G) for the 20 HPDs, for Q = 80% and 20%.

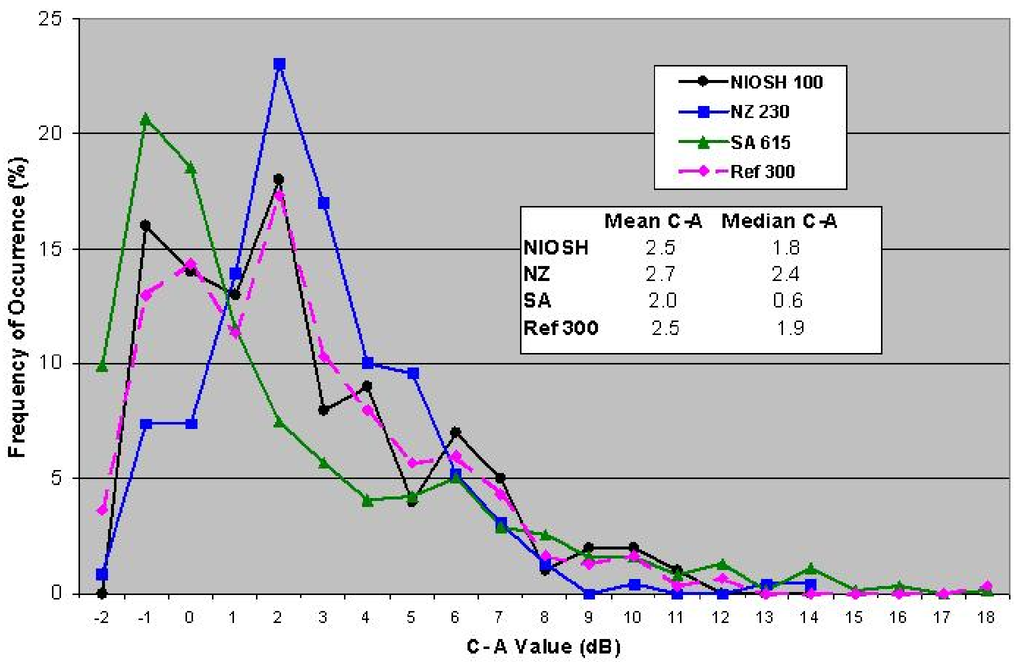


ANNEX B – THE NOISE DATABASES

The sources of the five noise databases that we utilized are discussed in Section 2.3. In this Annex we present a statistical description of those noises and the rationale for our choice of the Reference 300 (Ref300). The charts are presented in terms of the frequency of occurrence of noises as a function of the differences in their C-weighted minus A-weighted sound levels. This metric was selected since it provides an indicator of the spectral balance of the noises; higher C – A values are associated with noises that have more low-frequency content.

The distribution of the C – A values for the three sets of published industrial noises is presented in Figure B1 (Backshall, 2000; Kroes et al., 1975; McQueen at al., 1969). Notice that the three noise sets have different proportions of low- and high-frequency biased noises. However, the NIOSH and NZ noises have similar mean C – A values (2.5 and 2.7 dB) and similar median values as well (1.8 and 2.4 dB). The SA noises are more widely distributed and overall have more high-frequency content (low C-A values).

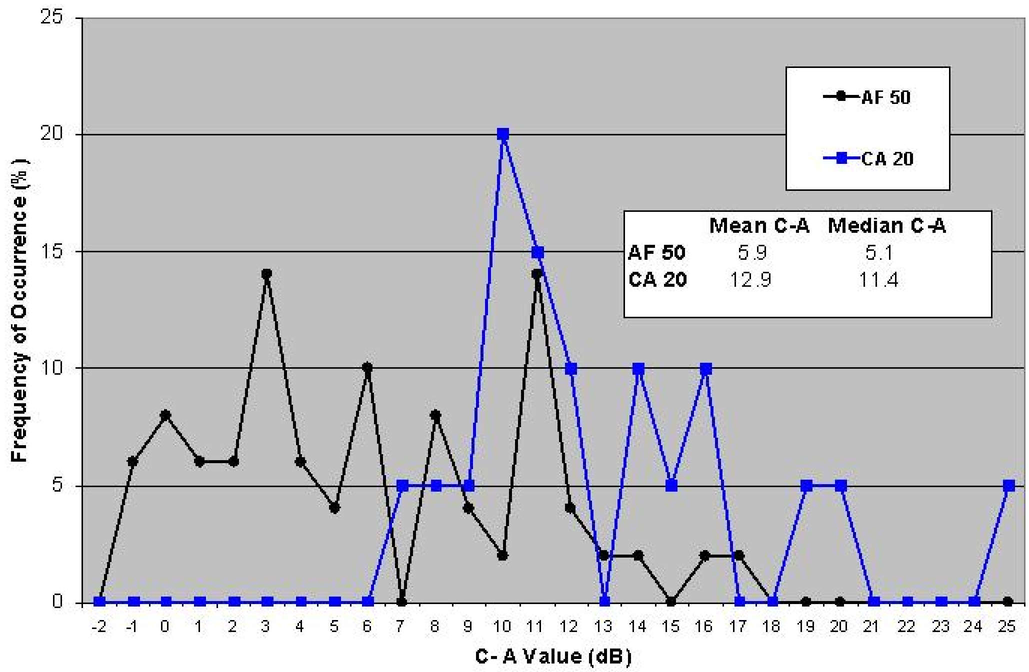
Table B1 – Distribution of the C – A values for the three industrial noise databases and for the Reference 300.



We decided that the best amalgam of the various databases could be obtained by creating a reference database of noises that would have equal representation from the three individual databases and also provide a larger number of noises against which to test the ratings. Thus all 100 NIOSH noises were included and on a quasi-random basis 100 each of the NZ and SA noises were also selected. The new database, Ref 300 is also shown in Figure B1.

Another well-known reference noise database is the Air Force 50 (Johnson and Nixon, 1974). In their analyses, Johnson and Nixon found that different outcomes were observed depending upon whether they used the NIOSH 100 or the Air Force 50 to assess HPD ratings. Since the Air Force 50 is targeted more towards a military rather than an industrial environment, we decided as did Johnson and Nixon, to utilize it separately. Additionally the first author had concerns that even the Air Force 50 did not capture enough

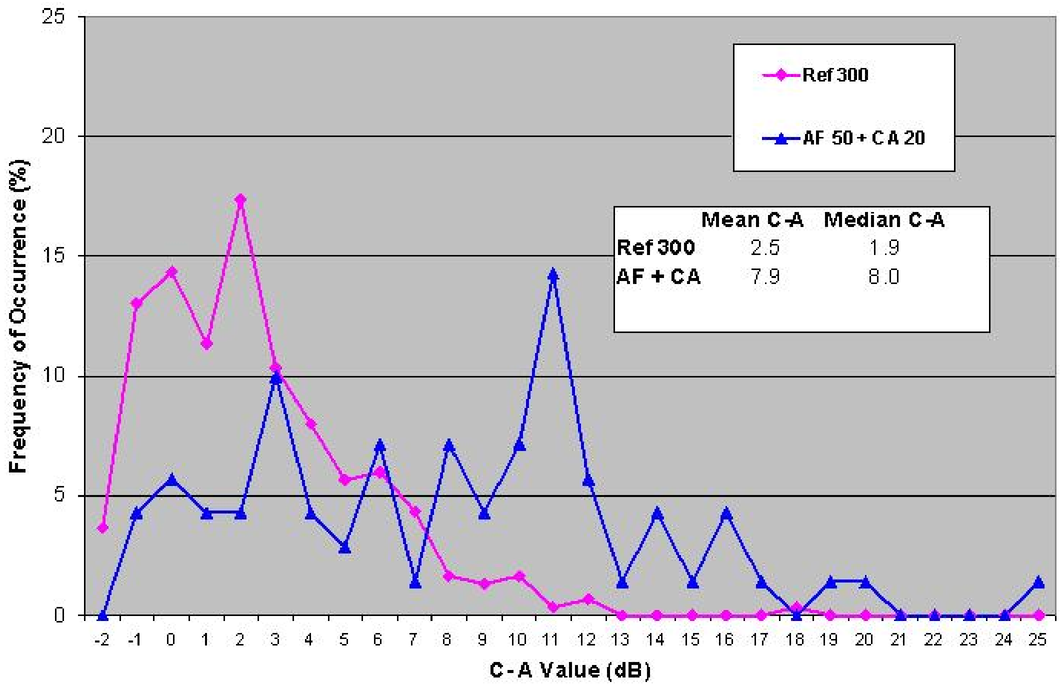
Figure B2 - Distribution of the C – A values for the Air Force and civil aviation noise databases.



noises with extreme low-frequency content. Therefore the Civil Aviation 20 was also included in our study. The two sets of spectra are illustrated in Figure B2. It is apparent that the two spectral collections have different distributions in terms of the percentage of high- and low-frequency noises they contain. Note the large difference between both their mean and median C-A values.

Finally, in Figure B3, the Ref 300 is compared to a combined set of the 70 Air Force and civil aviation noises. The data provide a clear indication of the broader distribution of spectral variation in the Air Force plus civil aviation noises. Analyses are reported separately for each of the databases in this report. The existence of the 70-noise database is one of the strong reasons for the inclusion of the NRS_G as secondary label information, since single-number A-based ratings do not provide sufficient predictive accuracy in such environments.

Figure B3 - Distribution of the C – A values for the Ref 300 vs. the Air Force and civil aviation noise database.



ANNEX C – COMPUTATIONAL DETAILS AND SAMPLE SPREADSHEET

In this Annex we provide equations defining the HPD ratings examined in this report. We also describe an Excel spreadsheet that is available from the authors which calculates the recommended ratings ($NRS_{A,80/20}$ and $NRS_{G,80/20}$) from either Method-B or Method-A data. Finally, the Matlab code used to perform all the calculations in this report is described; this is available from the first author.

Ratings: In the ensuing table of equations defining ratings examined in this report we use the following notation:

R_{ST} = REAT attenuation measured for an HPD on subject S in trial T. All attenuation data and noise spectra span the octaves from 125 Hz to 8 kHz; the frequency index is not shown explicitly.

pink = a noise spectrum with equal level in each octave-band.

$niosh_N$ = the N^{th} noise in the NIOSH100 database of industrial noise spectra.

$graph_N$ = the N^{th} noise in 170 spectra comprised of the NIOSH 100, the AF 50 and the CA 20 noises described in Annex B.

$C[X]$ = C-weighted level in dB of spectrum *X*. For example,
$C[niosh_N]$ = C-weighted level of N^{th} noise in NIOSH 100

$$= 10\log_{10}\left[\sum_{125-8k} 10^{[niosh_N + \mathbf{C}]/10}\right]$$ where $\mathbf{C}$ = C-weight filter response,

$\mathbf{C}$ = -0.2 at 125Hz, 0 at 250 to 1kHz, -0.2 at 2kHz, -0.8 at 4kHz, -3.0 at 8kHz.

$A[X]$ = A-weighted level in dB of spectrum *X*. For example,
$A[niosh_N]$ = A-weighted level of N^{th} noise in NIOSH 100

$$= 10\log_{10}\left[\sum_{125-8k} 10^{[niosh_N + \mathbf{A}]/10}\right]$$ where $\mathbf{A}$ = A-weight filter response,

$\mathbf{A}$ = -16.1 at 125Hz, -8.6 at 250Hz, -3.2 at 500Hz, 0 at 1khz, 1.2 at 2kHz, 1.0 at 4kHz, -1.2 at 8kHz.

$A[\,niosh_N - MN_T(R_{ST})\,]$ = Protected level A' in N^{th} NIOSH 100 noise for subject *S*

$$= 10\log_{10}\left[\sum_{125-8k} 10^{[niosh_N + \mathbf{A} - MN_T(R_{ST})]/10}\right]$$

$MN_J[X]$ = average (arithmetic mean) of *X* over all indices *J* shown in the subscript; e.g.,
$MN_N[A(niosh_N)]$ = mean A-weighted level of NIOSH 100 noises = $\sum_N A[\,niosh_N\,]\Big/100$.
$MN_T[R_{ST}]$ = subject-mean attenuation, i.e. the average across trials for each subject *S.*

$SD_J[X]$ = unbiased standard deviation of *X* over all indices *J* shown in the subscript (denominator is the degrees of freedom = the number of samples minus one); e.g.,
$SD_{ST}\,[R_{ST}]$ = standard deviation of REAT data used in computing the NRR

$$= \sum_{ST}\left[R_{ST} - MN_{ST}(R_{ST})\right]^2 \Big/ (30-1)$$ for 10 subjects X 3 trials.

Q = target protection rate (percent of population < 85 dBA) that a rating aims to achieve.

Z_Q = number of SD to subtract from the mean to achieve *Q* for normally distributed data.

NR = A-weighted noise reduction (A – A') defined by the rating or computed for a given noise *N* and subject *S* (NR_{NS}).

APV = assumed protection value, an attenuation response computed from the normal statistics by subtracting a multiple of the standard deviation from the mean attenuation.

$PC_Q[X]$ = value that *Q* percent of *X* are better than (lower in the case of protected level A', higher in the case of attenuation), computed directly rather than by normal statistics. Given an *m*-long set of values *X* ($x_1, x_2, \ldots x_m$) and the desired percentile *Q* (in the range 0 to 100) that *Q*% of *X* exceed, we compute $PC_Q[X]$ by interpolating as follows:

a) Sort *X* in increasing order and define $x_{m+1} = x_m$.
b) Let $(m+1)(1-Q/100) = i+r$, where *i* is an integer and *r* is the fractional remainder.
c) Then $PC_Q[X] = (1-r)x_i + rx_{i+1}$.

This percentile definition is the default used by SAS Institute software (SAS, 1995, definition #4) and matches Matlab's median function for representative cases that were tested.

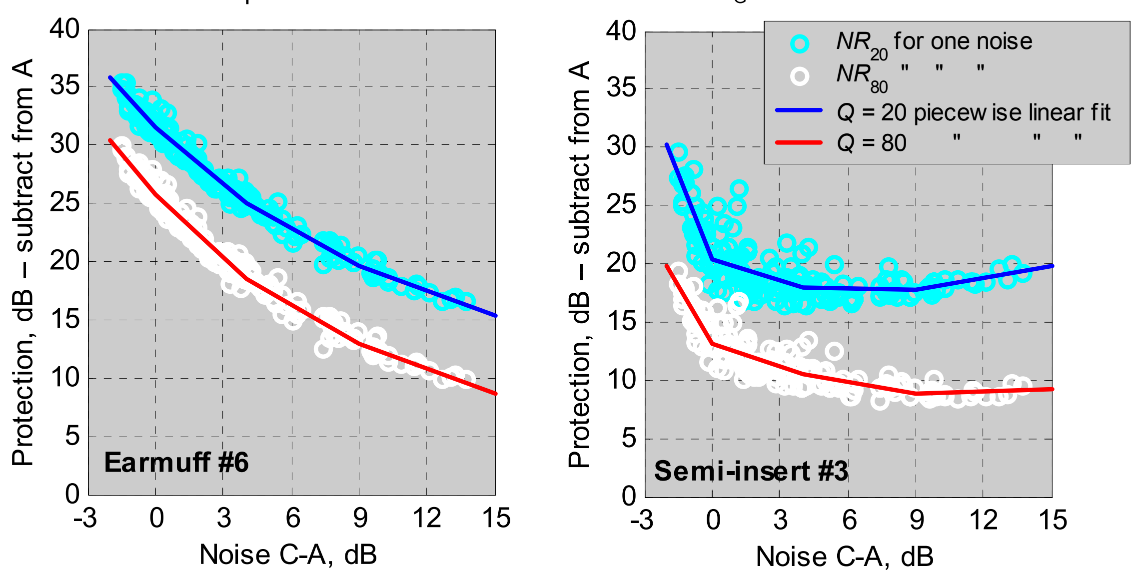
Table C1 – Definitions of ratings examined in this report

Rating	**Abbrev.**	**How computed**
Single number, C – A' type Estimate protected level A' by subtracting the rating value from the noise C-weighted level or TWA.		
Noise Reduction Rating, present EPA definition	NRR	$APV = MN_{ST}[R_{ST}] - 2*SD_{ST}[R_{ST}]$ $NRR = C[pink] - A[pink-APV] - 3$ *R* obtained using S3.19 (3.2 & 4k, 6.3 & 8k averaged)
SNR, ISO 4869-2[22]	SNR_Q	$APV_Q = MN_{ST}[R_{ST}] - Z_Q*SD_S[MN_T(R_{ST})]$ $SNR_Q = C[pink] - A[pink-APV_Q] + 0.5$
Single number, A – A' type Estimate protected level A' by subtracting the rating value from the noise A-weighted level or TWA.		
NRR applied to A-weighted noise	NRR–7	NRR minus 7 dB, the present government recommendation for how to adjust the NRR for use with A
	NRR_A	NRR_A = NRR – 2.5 (mean C – A for industrial noise)
NRR, applied to A, OSHA derated	$NRR_{A,OSHA}$	$NRR_{A,OSHA} = (NRR - 7) / 2$
NRR, applied to A, NIOSH derated	$NRR_{A,NIOSH}$	$NRR_{A,NIOSH} = NRR*D - 7$ *D = 0.75* earmuff, *0.5* foam plugs, *0.3* other plug
Noise Reduction Factor applied to A	$NRF_{A,Q}$	$APV_Q = MN_{ST}[R_{ST}] - Z_Q*SD_S[MN_T(R_{ST})]$ $NRF_{A,Q} = C[pink] - A[pink-APV_Q] - 3 = NRR_{SF} + 1.5$
Modified CSA Z94.2	Class	$SNR_{84} \geq 8$ Class 1 use $NR = 5$ ≥ 12 Class 2 $NR = 10$ ≥ 16 Class 3 $NR = 15$ ≥ 20 Class 4 $NR = 20$
Noise Reduction Statistics, applied to A	$NRS_{A,Q}$	$NR_{NS} = A[niosh_N] - A[niosh_N - MN_T(R_{ST})]$ for all subjects *S* and noises *N* $NRS_{A,Q} = MN_{NS}[NR_{NS}] - Z_Q*SD_{NS}[NR_{NS}]$
Noise Reduction Percentile, applied to A	$NRP_{A,Q}$	NR_{NS} same as for NRS_A $NRP_{A,Q} = PC_Q[NR_{NS}]$, percentile directly computed

[22] The 0.5 dB adjustment is needed because the specified pink spectrum in ISO 4869-2 is assigned a level C = 100 irrespective of whether the A' level is computed from 63 or 125 Hz. When computed from 125 Hz as we do, C = 99.5, so 0.5 dB must be added to our value to match ISO 4869-2. The target protection rate *Q* is referred to as α in ISO 4869-2.

Multi-number		
Octave-band, normal statistics	OBN_Q	$APV_Q = MN_{ST}[R_{ST}] – Z_Q{*}SD_S[MN_T(R_{ST})]$ A’ = $A[N–APV_Q]$ when HPD is used in noise spectrum N
Noise Reduction Statistics, graphical	$NRS_{G,Q}$	$NR_{NS} = A[graph_N] – A[graph_N – MN_T(R_{ST})]$ for all subjects S and noises N. $NR_{N,Q} = MN_S[NR_{NS}] – Z_Q{*}SD_S[NR_{NS}]$ for each noise N. $NR_{N,Q}$ is then subdivided into five C – A ranges bounded by the values –2, 0, 4, 9, and 15 dB. Within each range a linear regression is performed on $NR_{N,Q}$ as a function of the noise C – A value and the values of NR_Q at the C – A bounds are calculated from the best-fit line’s slope and intercept. The resulting NR_Q values for the adjacent line segments are then averaged to yield five NR_Q values defining the graph for each Q value (see Figure C1). To apply the NRS_G, one looks up the protection NR corresponding to the noise environment’s C – A value on the graph. Alternatively, NR may be found by linear interpolation of the two nearest points on the graph. For example, if the noise C – A value is 7 dB and the $NRS_{G,80}$ values at C – A = 4 is 16 and at C – A = 9 is 10, then the correct NR value to use would be 16+(7–4)*(10–16)/(9–4) = 12.4 dB. In many situations, the estimate of the true C – A value in the environment may have enough imprecision that simply using the NR_Q values at the nearest of the five C – A values is sufficiently accurate.

Figure C1 – Illustration of piecewise linear fit used to calculate NRS_G.



Note that the C – A values for the 170 noises that define the NRS_G are computed over the frequencies from 125 Hz to 8 kHz. However, in applying the NRS_G, C – A would most likely be measured by a sound level meter or C-weighting capable dosimeter that would include in its measurements the noise energy below 125 Hz. Therefore, the resulting computed C – A values are lower than the true (measured) C – A values for each noise, so the points defining the graph (the circular marks in Figure C1) are shifted to the left from their true positions. Since $NRS_{G,80}$ tends to decrease for increasing C – A, this would appear to lead to an understatement of protection by NRS_G at high C – A. In actuality, this is beneficial since REAT tends to overstate attenuation at low frequencies due to the effects of physiological noise masking (Gauger, 2003); the two effects counterbalance each other to a degree.

Spreadsheet: An Excel workbook (NRSa & g, WG11 ratings.xls) has been created to calculate the recommended ratings, NRS_A and NRS_G. Below is a brief description of the structure and use of the spreadsheet. First, the "<u>input & calcs</u>" sheet shown in Figure C2 is described.

1) All input cells on the spreadsheet are shaded light blue; these are the only cells into which data should be entered or that should be modified.
2) The spreadsheet is structured presuming use of REAT data with up to twenty subjects with two trials per subject. This is typed or pasted into cells D9:J48 – any unneeded input cells (e.g., if data for an earmuff with only 10 subjects is being entered) should be left blank.
3) A description of the device tested and any other pertinent information about the test (laboratory, date, test index number, etc.) should be entered into cell C5. This information is mirrored on the output page of the spreadsheet.
4) Either "B" or "A" should be typed into cell I6 to indicate which ANSI S12.6-1997 method was used to measure the REAT data. This is used to set derating parameters used in cells I4:J4. The target protection rates *Q* defining the upper and lower values of the ratings are specified in cells I2:J2. These cells are fixed and not user definable.
5) The subject-mean attenuation values are computed in cells M9:T29 and each subject's A – A' value in pink noise is shown in column M. To the right of these cells, beginning with column Y, the individual subject-mean noise reductions for each of the 170 noises (one column per spectrum) in the NIOSH 100, AF 50 and CA 20 are calculated using Excel array formulas. Array formulas eliminate the need to have cells containing intermediate results in performing calculations such as are used by the octave-band method; see Excel's documentation for more information on array formulas.
6) The list of subjects in column L automatically updates to reflect the number of subjects for whom data was input. All the formulas in columns M:GL reference the corresponding cell in column L so that results are only calculated for the actual number of subjects. This allows the spreadsheet to easily be expanded to use data for more than 20 subjects (e.g., if pooling data from multiple tests). See the instructions provided in the spreadsheet.
7) At the top of columns L:T are two charts, one showing a histogram of the A – A' value in pink noise across subjects (reflecting the degree of non-normality of the data) and the other showing the relative variation in noise reduction due to across-subject and across-spectrum effects.
8) The 170 spectra used in columns Y:GL are sorted by increasing C – A value; an "N" in row 7 indicates if the spectrum is one of the NIOSH 100. In cells Y33:GL44 the C – A and $NR_{Q,S}$ values needed to compute the NRS_G graph are computed.

Figure C2 - Input sheet of NRS_A / NRS_G Excel workbook.

INSTRUCTIONS: Input cells are shaded light blue. Enter a description of the device, lab, etc. into the shaded cell below and then the REAT data into the cells further down. To change the calculation parameters to use Method-A or Method-B data, change the contents of cell I6. If you have fewer than 20 subjects leave the unneeded cells blanks; if you have more than 20 subjects follow the instructions in cell L33. All cells that are not shaded are protected against inadvertant change.

Percent protected (upper & lower limits and derating multipliers)

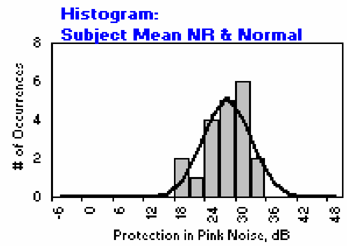
80%	20%
-0.84sd	0.84sd
x 1.	x 1.

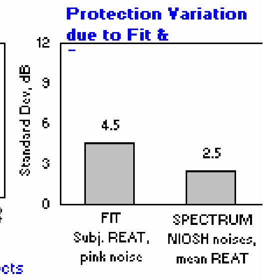
S12.6 method (type "**B**" or "**A**")

B

Device, manufacturer, lab, date, test #: **Foam earplug #3, 20 HPD database EARCAL test ID# 321003**

Subject	Trial	Trial NR	125	250	500	1k	2k	4k	8k
			Frequency						
1	1	20	14.	18.	13.	21.	19.	35.	33.
	2	13	16.	10.	9.	8.	32.	38.	32.
2	1	33	25.	26.	28.	30.	35.	41.	50.
	2	35	26.	29.	34.	35.	35.	38.	37.
3	1	22	19.	16.	18.	17.	29.	40.	35.
	2	21	23.	19.	16.	16.	33.	40.	33.
4	1	26	14.	20.	23.	22.	30.	37.	35.
	2	26	22.	19.	22.	23.	28.	38.	30.
5	1	32	24.	26.	28.	30.	33.	41.	42.
	2	24	21.	18.	20.	20.	34.	41.	41.
6	1	30	33.	25.	26.	27.	35.	38.	36.
	2	26	27.	24.	19.	24.	37.	38.	34.
7	1	28	23.	23.	27.	24.	30.	42.	43.
	2	30	19.	23.	25.	29.	30.	43.	44.
8	1	34	33.	37.	43.	45.	29.	46.	46.
	2	34	37.	39.	40.	42.	29.	46.	43.
9	1	29	23.	28.	23.	26.	36.	45.	44.
	2	26	19.	22.	22.	21.	33.	43.	39.
10	1	22	16.	14.	18.	18.	29.	36.	42.
	2	26	27.	21.	22.	22.	29.	41.	45.
11	1	30	26.	30.	32.	37.	25.	38.	38.
	2	31	21.	24.	31.	31.	28.	38.	36.
12	1	27	23.	24.	20.	24.	37.	44.	36.
	2	23	17.	18.	18.	18.	34.	38.	40.
13	1	23	22.	19.	23.	17.	37.	42.	41.
	2	31	25.	28.	26.	28.	38.	40.	38.
14	1	17	14.	8.	12.	19.	15.	29.	29.
	2	28	20.	23.	25.	24.	29.	43.	44.
15	1	28	23.	24.	27.	23.	29.	41.	42.
	2	28	17.	20.	28.	26.	28.	38.	40.
16	1	32	27.	30.	29.	28.	33.	40.	39.
	2	30	23.	27.	27.	26.	30.	45.	42.
17	1	20	16.	15.	15.	18.	22.	27.	31.
	2	17	12.	15.	9.	17.	25.	31.	28.
18	1	28	18.	20.	24.	25.	33.	44.	44.
	2	24	18.	17.	21.	19.	29.	41.	44.
19	1	31	28.	27.	26.	28.	36.	42.	45.
	2	29	24.	24.	21.	30.	38.	43.	44.
20	1	21	7.	7.	20.	24.	32.	36.	26.
	2	25	19.	16.	20.	22.	36.	44.	42.





Subject mean attenuatio 20 subjects

Subject	Subject NR	125	250	500	1k	2k	4k	8k
		Frequency						
1	18	15	14	11	15	26	37	33
2	34	26	28	31	33	35	40	44
3	21	21	18	17	17	31	40	34
4	26	18	20	23	23	29	38	33
5	29	23	22	24	25	34	41	42
6	29	30	25	23	26	36	38	35
7	29	21	23	26	27	30	43	44
8	34	35	38	42	44	29	46	45
9	28	21	25	23	24	35	44	42
10	24	22	18	20	20	29	39	44
11	31	24	27	32	34	27	38	37
12	25	20	21	19	21	36	41	38
13	28	24	24	25	23	38	41	40
14	23	17	16	19	22	22	36	37
15	28	20	22	28	25	29	40	41
16	31	25	29	28	27	32	43	41
17	19	14	15	12	18	24	29	30
18	26	18	19	23	22	31	43	44
19	31	26	26	24	29	37	43	45
20	24	13	12	20	23	34	40	34
-	-	-	-	-	-	-	-	-

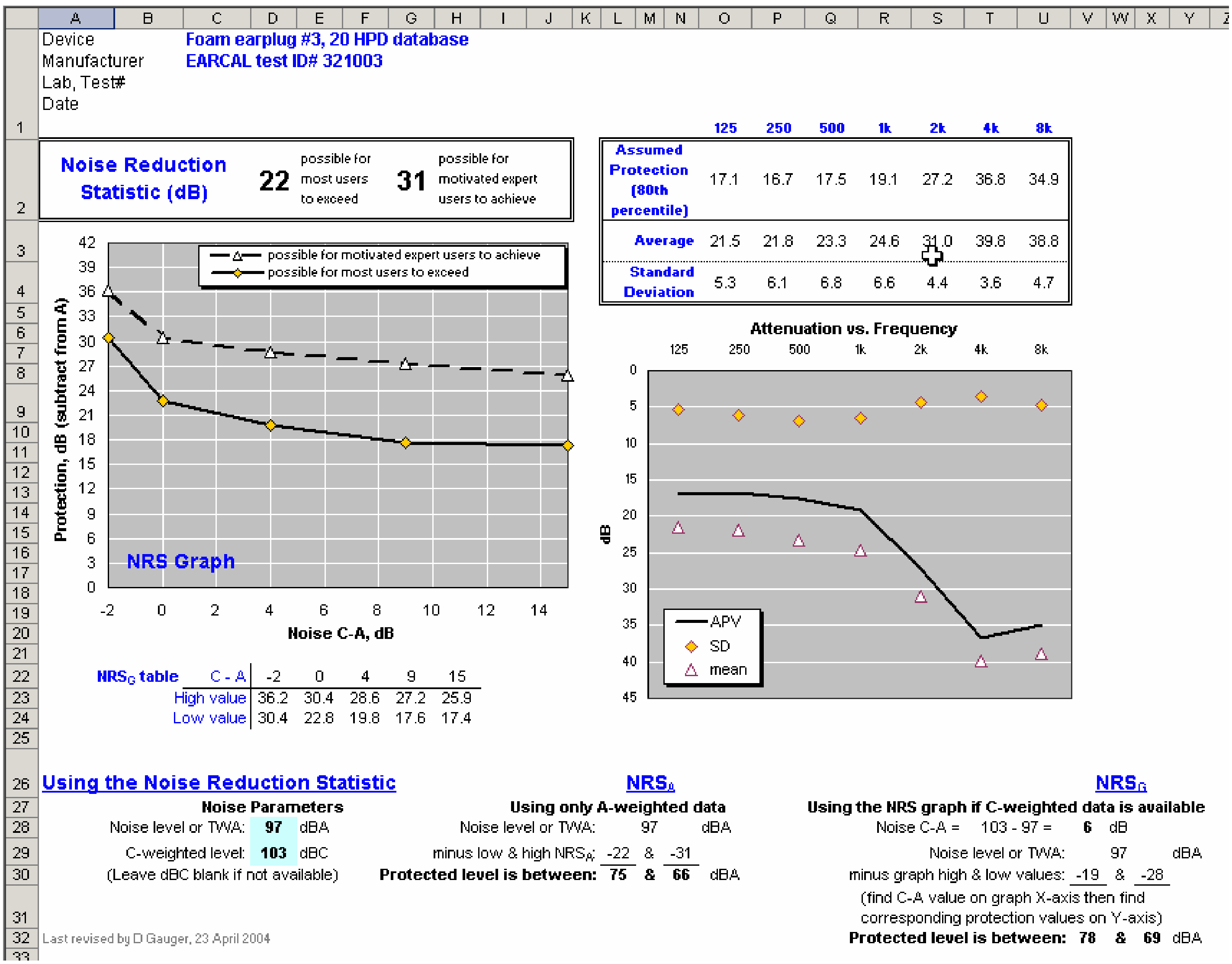
To add rows if you have more than 20 subjects:
1) Unprotect the worksheet (Tools/Protection... menu). No password is required.
2) Insert the number of rows you need for the additional subjects, starting with row For example, for 5 subjects, highlight rows 11:20 (10 rows) and insert. You should always insert an even number of rows (two trials per subject required).
3) Select row 10, copy, then select all of the newly-inserted rows and the first pre-existing row below the inserted ones and paste. In the above example copy row 10 into rows 11:21.
4) The subject and trial columns (A:B and L) will be updated for the number of subjects you have. Enter the REAT data into the shaded cells.

Next, turning to the "ratings" sheet shown in Figure C3.

9) The NRS_A values are computed using array formulas in cells D2 and G2. These refer back to the NR_{NS} results as well as the Z_Q and, if the data are Method A, derating values on the input sheet. A range of cells on the input sheet named "niosh" (cells Y7:GL7) are tested to see if the noise spectrum is one of the NIOSH 100; only these spectra are used to compute the NRS_A values.
10) The octave band normal statistics and APV corresponding to Q = 80% (or the value specified in input sheet cell I2) are calculated in cells O2:U4 and plotted in a corresponding figure in those columns.

11) The Excel TREND function is used in cells D23:H24 to compute the end points of each line segment in the NRS_G graph, referring back to the corresponding subsets of C – A and $NR_{Q,S}$ values on the input sheet. The NRS_G graph appears above the table.
12) Starting at row 27 the spreadsheet is set up to provide an example of the application of the rating values. Noise parameters (A- and C-weighted levels) may be entered in cells D28:D29. Cells L28:N30 illustrate the use of NRS_A and cells V28:X32 illustrate the use of NRS_G.

Figure C3 - Ratings sheet of NRS_A / NRS_G Excel workbook.



Matlab toolbox and workspace: All calculations done in this report were performed using a collection of functions written in Matlab collected together into a toolbox called HPrate. This has been successfully used under both release 12 and release 13. Some functions compute rating values, others generate or compute metrics upon a Waugh analysis population while others perform basic functions such as computing A-weighted levels or statistics on attenuation data. A listing of the HPrate/Contents.m file describing each function follows; it gives a one-line description of each function as well as the format of data used by the functions. The code is well commented and should be reasonably easy to follow for those conversant in Matlab's technical computing language. This toolbox, along with a workspace containing all the data used in this report and scripts used to generate most of the figures and tables in this report are available from the first author upon request.

```
HPrate/Contents.m
% COLLECTION OF FUNCTIONS FOR MANIPULATING VARIABLES CONTAINING HEARING
% PROTECTOR ATTENUATION DATA - SPECIFICALLY TO DISPLAY THE ATTENUATION,
% CALCULATE RATINGS DISTILLED FROM IT AND ALSO TO CALCULATE PERFORAMNCE METRICS
% ON THESE RATINGS USING THE METHOD ORIGINALLY PROPOSED BY DICK WAUGH.
%
%ALL FUNCTIONS ARE LISTED BELOW.  IF THE NAME IS IN PARENTHESIS THEN THAT
%FUNCTION MAY NOT YET EXIST OR BE FULLY FUNCTIONAL (note: a .m file
%may already exist containing terse notes about how to approach the code). IF
%THE NAME IS FOLLOWED BY A QUESTION MARK, IT WORKS BUT IMPROVEMENTS ARE NEEDED.
%
%RATINGS ================================================================
%        A set of one or more numbers used to summarize the attenuation
%        of an HPD across a set of subjects/trials from which Ap (protected
%        level) can be estimated from some summary info about a noise.
%        NOTE: in help comments for functions, < > denote optional inputs.
%    OBN        7-number "octave band normal" rating (mean-N*sds in each band).
%    OBP        7-number "octave band percentile" rating from each band subject mean
%    NRFa       Generalized NRRsf.  Definition: C-Ap-K in pink.  Use: Ap=A-NRFa
%    NRFc       Generalized Botsford rating. Def: C-Ap-K in pink. Use: Ap=C-NRFc.
%    NRR        Present EPA NRR on S3.19 data if exists. Def:C-Ap-3,pink Use:Ap=C-NRR
%    NRRm7      Current EPA NRR applied to A using official 7dB correction & optional derating
%    NRRa       Current EPA NRR applied to A with 7dB correction & optional derating
%    NRSa       2 number A-Ap rating from normal stats on all REAT subj. means x NIOSH
%    NRPa       2 number A-Ap rating = percentiles on all REAT subject means x NIOSH
%    NRPat      2 number A-Ap rating = percentiles on all REAT trials x NIOSH
%    NRSc       2 number C-Ap rating from normal stats on all REAT subj. means x NIOSH
%    NRPc       2 number C-Ap rating = percentiles on all REAT subject means x NIOSH
%    NRPct      2 number C-Ap rating = percentiles on all REAT trials x NIOSH
%    NRSg       Graph of noise reduction from normal stats vs C-A proposed for 2ndry label
%    NRPg       C-A graph percentiles graph rating proposed for secondary label
%    SNR        ISO4869-2 SNR84 excluding the 63 Hz band = NRFc(reat,<noises>,0,-0.5 or 0)
%    NRRsf      NHCA Task Force NRR(SF).   Def: SNR-5   Use: Ap=A-NRRsf
%    Z94        Generalized Z94-like grading system where grade assigned based on SNR
%   (HML)       ISO4869-2 hi/mid/lo 3-number rating excluding 63Hz band
%    FCob?      Simulate octave band "fit check" to determine protected level
```

```
%   (FCsf)      Simulate fit-check that estimates INR from 1 or 2 frequencies
%    Ratings    Script to build/graph a table of rating values for several HPDs
% The following 3 are other measures of attenuation data but not actually ratings:
%    sdspec     A-A' std dev across noise set (default=niosh) for mean HPD/group REAT
%    sdfits     HPD or groups A-A' std dev across subjects in pink or an input noise
%    sdfitt     HPD or groups A-A' std dev across trials in pink or an input noise
%
%METRICS ================================================================
%         A measure of how accurately some rating performs; ie, how well it
%         assures that a population is protected if the rating is trusted. This
%         is determined by running the function WAUGH to get the Ap for the
%         population then doing statistics on that population. Two globals
%         control this.
%           modeWaugh = see WAUGH help for explanation
%           modeMetric = 'waugh'...metric calculated on whole population - default
%                      = 'murphy'...metric a la B Murphy (calculate stat in each noise
%                                  then return mean & std err of mean across noises)
%    waugh      Ap for all subject/trial X noise combos after adj. noise level per rating
%    histAp     Plots a histogram of one or more population Ap matrices from WAUGH
%    Pprot      Calcs % protected (Ap < crtrn) from population returned by WAUGH
%    Pidl       Calcs % ideal (ovrpro <= Ap < crtrn) from population returned by WAUGH
%    Pover      Calcs % overprotected (Ap < ovrpro) from population returned by WAUGH
%    Pundr      Calcs % underprotected (Ap >= crtrn) from WAUGH population
%    ApSD       Calcs Ap standard deviation from population Ap returned by WAUGH
%    Apundr     Calculates Ap = average dose for underprotected part of WAUGH population
%    ApMN       Calcs mean from population Ap double or cell vector returned by WAUGH
%    Ap1        Calculates Ap value that 1% of the Waugh population exceed
%    Ap2        Calculates Ap value that 2% of the Waugh population exceed
%    Ap5        Calculates Ap value that 5% of the Waugh population exceed
%    SDhPprot   Std deviation & worst case % prot, Apundr for a rating over HPD group
%    SDnPprot   Std deviation & worst case % prot, Apundr over noises for HPD or group
% The following are not metrics as defined above but belong here more than anywhere else
%    rsds       Standard deviation of rating by bootstrapping to sim subjects
%    msdn       Standard deviation of metric by resampling noises used by WAUGH
%
%BASIC UTILITY FUNCTIONS ================================================
%    HPDs       Script to list all REAT structures with labels & included method fields
%    group      Script for building a group of HPDs for consolidated analysis
%    ideal      Creates ideal normally distributed REAT with target mean & s.d.
%    inr        Graphs individual subj noise reductions & fits uni/bimodal model
%    atten      Displays a table of the values in a REAT variable
%    atteng     Graphs the values of a REAT variable along with MN and SDS
%    attenm     Graphs the mean +/- N s.d. of a set of REAT variables
%    attens     Graphs all values of up to three REAT variables.
%    awt        Returns A-weighted levels for set of noises or noise/hpd combos
%    cwt        Returns C-weighted levels for set of noises or noise/hpd combos
%    consol     Consolidates Hx1 cell vector of NxS Waugh pops into one NxSH pop
```

```
%    decimals   Rounds double to ? decimal places - default=0, persistently changeable
%    flatten    Turns REAT Subj x Trial x Bands double into an ST x B double
%    mn         Returns the unrounded mean value of a REAT double
%    mns        Returns the subject mean (across trials) for a REAT double
%    normdist   Converts SD / percentile to percentile / SD, with 100% = -inf SD
%    pcntle     Returns value that given percent of input data exceed.
%    sds        Returns unrounded standard deviation across subjects of a REAT
%    sdt        Returns unrounded standard deviation across trials for a REAT
%    stparse    Parses 2D (STxB) REAT table into an SxTxB double (for pasting in data)
%    r2m        Subtracts (or adds) physiological noise masking effect from REAT (to MIRE) data
%
%VARIABLE TYPES ==========================================================
%   HPD attenuation test data is stored in structures named in the form
%   BBddd where BB denotes the database (eg, TP for "20 protector") and ddd
%   is a short code for the particular device. Each HPD contains a separate
%   field giving data from a given test method/lab; the data is organized as
%   an S x T x 7 table (subjects x trials x #bands). A ".la" field provides
%   a description of the source of the data (NOTE: this field should include
%   words "muff", "plug", or "foam" for NRR derating to work).  The method
%   fieldnames used must be one character.  If the data field is sized 2x1x7
%   then it was created by IDEAL to match published mean and SD values since
%   individual subject data was not available.  Method fieldnames used include:
%               B = S12.6 Method-B subject fit REAT measured at EARCAL
%               S = S3.19 experimenter fit REAT, 3.2k avg'd into 4k, 6.3k into 8k
%                   as reported by the manufacturer
%               I = ISO 4869-1 experimenter supervised fit REAT as reported
%                   by the manufacturer
%               E = S3.19 experimenter fit REAT measured at EARCAL or NIOSH
%                   if S field data is not from EARCAL (ie, not EAR product)
%               F = field (real world) REAT from NIOSH studies
%               V = Method-B measured at VaTech in VE* dataset
%               M = MIRE, subject fitted, measured at Bose
%               R = REAT passive + MIRE active
%GROUPS Cell arrays containing HPD data and the names of different REAT
%       variables along with their data and method fieldnames; use to feed rating
%       and metric calculation. Built using the script "group"--see it's help file.
%       Generally named in form gBB*.
%NOISES Doubles containing N x #bands sets of noise spectra.
%       pink = pink noise with A=100
%       niosh = the 100 "NIOSH noises"
%       sa615 = the 615 spectrum South Australian noise database
%       nzgt85 = the 230 spectra from Backshall's New Zealand database > 85 dBA
%       ref300 = a 300 spectrum subset of NIOSH+SA+NZ (global)
%       noisesInd = concatenation of industrial noises: niosh,sa615,nzgt85
%       af50 = the Johnson/Nixon 50 spectra set of Air Force noises
%       ca20 = D Gauger collection of 20 civil aviation spectra
%       noisesAvi = concatenation of aviation noises: af50,ca20
```

```
%       armor = 14 spectra from military armored vehicles
%GLOBAL VARIABLES used by various functions
% bands  = cell vector of strings for labeling frequency bands. Checked by many
%          functions to see if the calculation should include 63 Hz.
% crtrn  = criterion Ap -- max "safe" level
% exchrt = exchange rate -- allowable dB increase for a time-halving
% ovrpro = Ap level below which a user is considered overprotected
% pink, niosh, ref300, noisesInd = noise spectra defined above
% modeProgress = set to 1 to turn on calculation progress displays for
%                computations that take awhile (e.g. running RSDS or MSDN,
%                particularly on HPD groups with NRP ratings). 0 disables
%                progress displays.
% modeMetric = determines how metrics calculated on waugh output population Ap
%               See METRICS above for an explanation of these
% modeWaugh  = determines how waugh handles trials in generating population Ap
%               See HELP WAUGH for an explanation
% nrpcntls = 2x1 double giving default percentile values returned by ratings NRP*
% thisHPDtext = patch added Dec03 - char variable used by Waugh & Ratings
%               to pass HPD's .la field to ratings such as NRR (used in derating).
%
%REVISION HISTORY
%   Originally written by DGauger, Bose Corp, Feb-Apr 2003
%   Contents.m revised 14 Dec 03 per discussions with Berger, Franks, Murphy
%   Extended Dec-Jan04: inr, histAp, NRR derating, NRRa, SD*Pprot, rsds, msdn,
%                       normdist, NRSa/c & many misc improvements to other routines
%
```