

# **A review of the history, development and application of auditory weighting functions in humans and marine mammals**

[Dorian S. Houser](#), [William Yost](#), [Robert Burkard](#), [James J. Finneran](#), [Colleen Reichmuth](#), and [Jason Mulsow](#)

Citation: [The Journal of the Acoustical Society of America](#) **141**, 1371 (2017); doi: 10.1121/1.4976086

View online: <http://dx.doi.org/10.1121/1.4976086>

View Table of Contents: <http://asa.scitation.org/toc/jas/141/3>

Published by the [Acoustical Society of America](#)

---

---

# A review of the history, development and application of auditory weighting functions in humans and marine mammals

Dorian S. Houser,<sup>1</sup> William Yost,<sup>2</sup> Robert Burkard,<sup>3</sup> James J. Finneran,<sup>4</sup>  
Colleen Reichmuth,<sup>5</sup> and Jason Mulsow<sup>1</sup>

<sup>1</sup>National Marine Mammal Foundation, 2240 Shelter Island Drive, Suite 200, San Diego, California 92106, USA

<sup>2</sup>Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287, USA

<sup>3</sup>Department of Rehabilitation Science, University at Buffalo, 510 Kimball Tower, Buffalo, New York 14214, USA

<sup>4</sup>United States Navy Marine Mammal Program, Space and Naval Warfare Systems Center Pacific, Code 71510, 53560 Hull Street, San Diego, California 92152, USA

<sup>5</sup>Institute of Marine Sciences, Long Marine Laboratory, University of California Santa Cruz, Santa Cruz, California 95060, USA

(Received 14 September 2016; revised 29 December 2016; accepted 25 January 2017; published online 2 March 2017)

This document reviews the history, development, and use of auditory weighting functions for noise impact assessment in humans and marine mammals. Advances from the modern era of electroacoustics, psychophysical studies of loudness, and other related hearing studies are reviewed with respect to the development and application of human auditory weighting functions, particularly A-weighting. The use of auditory weighting functions to assess the effects of environmental noise on humans—such as hearing damage-risk criteria—are presented, as well as lower-level effects such as annoyance and masking. The article also reviews marine mammal auditory weighting functions, the development of which has been fundamentally directed by the objective of predicting and preventing noise-induced hearing loss. Compared to the development of human auditory weighting functions, the development of marine mammal auditory weighting functions have faced additional challenges, including a large number of species that must be considered, a lack of audiometric information on most species, and small sample sizes for nearly all species for which auditory data are available. The review concludes with research recommendations to address data gaps and assumptions underlying marine mammal auditory weighting function design and application.

© 2017 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4976086>]

[WWA]

Pages: 1371–1413

## I. INTRODUCTION

### A. A review

A variety of auditory weighting functions have been developed for both humans and non-human animals to measure sound in a biologically relevant manner. Auditory weighting functions have a storied history in the measurement of sound related to human activity, especially workforce safety and hearing loss, but also to effects such as annoyance and masking. Weighting functions have more recently gained momentum in the marine mammal community due to a desire to better predict the auditory effects of anthropogenic sound on marine mammals. The developmental pathways and implementations of auditory weighting functions in humans and marine mammals have not followed parallel paths, driven partly by the available data to inform the weighting functions and partly by the desired application of the functions. Weighting functions are used to assess a variety of auditory effects on humans (e.g., annoyance, masking, hearing loss); while the effort to develop auditory weighting functions in marine mammals has been driven almost entirely by a desire to more accurately predict and prevent noise-induced hearing loss (NIHL).

In this review, the history and development of auditory weighting functions are described both for humans and

marine mammals. Terms used in this document are either defined in the text, are defined in ANSA/ASA S1.1, *Acoustic Terminology*, or are discussed in the [Appendix](#). The rationale for certain approaches to the development and implementation of auditory weighting functions is provided, as well as various assumptions that have gone into the process. Where steps in the development and application of weighting functions could not be determined from the available literature, which is almost exclusively for human applications, this is also identified. The purpose of the review is to provide a better understanding of auditory weighting functions, their usefulness in predicting effects on hearing, and where issues with weighting function development and application exist. The review concludes with recommendations for research that might better inform the development of weighting functions and their application, specifically as related to marine mammal environmental issues.

### B. Auditory weighting functions

Sound has the physical properties of magnitude, frequency, and time, but animal hearing is not equally sensitive to acoustic magnitude at all frequencies. *Auditory weighting functions* transform sound measurements to take into account

the frequency-dependent aspects of auditory sensitivity. They are mathematical functions used to emphasize frequencies where animals (human and non-human) are more sensitive and de-emphasize frequencies where animals are less sensitive. Weighting functions may be thought of as frequency-dependent filters that are applied to sound before a single, weighted sound level is calculated [e.g., the sound pressure level (SPL)]. The filter shapes are normally “band-pass” in nature; that is, the function amplitude resembles an inverted “U” when plotted versus frequency. The weighting function amplitude is approximately flat within a limited range of frequencies, called the “pass-band,” and declines at frequencies below and above the pass-band.

Auditory weighting functions have been used to provide measures of the potential adverse effect (e.g., hearing loss) of sound on humans and, to a lesser extent, other non-human animals. In many cases, the frequency-dependent effects of a sound are used in conjunction with auditory weighting functions to derive a *weighted impact threshold*. The sound received by the animal is “weighted” by adding the auditory weighting function amplitude [in decibels (dB)] to the noise spectral amplitude (in dB) at each frequency (note that weighting functions often have negative amplitude values).

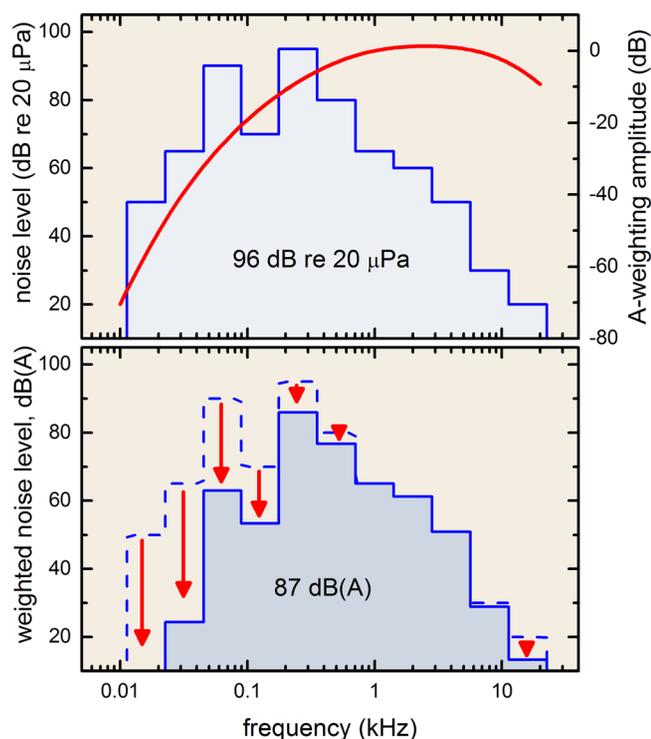


FIG. 1. (Color online) (Top) The blue line shows a hypothetical, octave-band sound pressure spectrum in air, with a total sound pressure level (integrated over all octave bands) of 96 dB re 20  $\mu\text{Pa}$ . The red line shows the human A-weighting function amplitude (an auditory weighting function described in this report). (Bottom) To determine the weighted exposure level, the A-weighting amplitude at each frequency is added to the sound pressure level at each frequency (red arrows). The weighted spectrum has lower amplitude at the frequencies where the A-weighting function amplitudes are negative. The values from  $\sim 1$ –4 kHz do not change significantly, since the weighting function is flat (i.e., the weights are near zero). The weighted SPL is calculated by integrating the weighted spectrum across all octave bands; the result is 87 dBA, meaning a sound pressure level of 87 dB re 20  $\mu\text{Pa}$  after applying the human A-weighting function.

The weighted noise spectral density levels are then integrated across frequency to yield a single, weighted sound level. The weighted sound level is compared to the weighted impact threshold; if the weighted sound level exceeds the weighted impact threshold, then an adverse impact (e.g., an estimated hearing loss) is assumed to occur (see Fig. 1).

Comparing different weighting functions is only meaningful if the weighted thresholds are also taken into account. For this reason, it is helpful to also consider the *acoustic exposure function*, defined as the difference between the weighted impact threshold and the weighting function amplitude at each frequency (Fig. 2). If the weighting function is normalized so the peak amplitude equals 0 dB, the exposure function is obtained by inverting the weighting function and setting the minimum value to the weighted impact threshold. The exposure function provides a visual indication of the manner in which the impact thresholds vary with frequency, and may be directly compared to temporary or permanent NIHL onset values at different frequencies.

Auditory weighting functions as described in Fig. 1 are a common means of transforming sound measurements relevant to humans, and several types of weighting functions have been developed for applications to human hearing. By comparison, the application of auditory weighting functions to non-human animals is limited; they have received the greatest use in a limited scope of marine mammal auditory effects. However, in all cases, the objective of applying an auditory weighting function to sound measurements is to

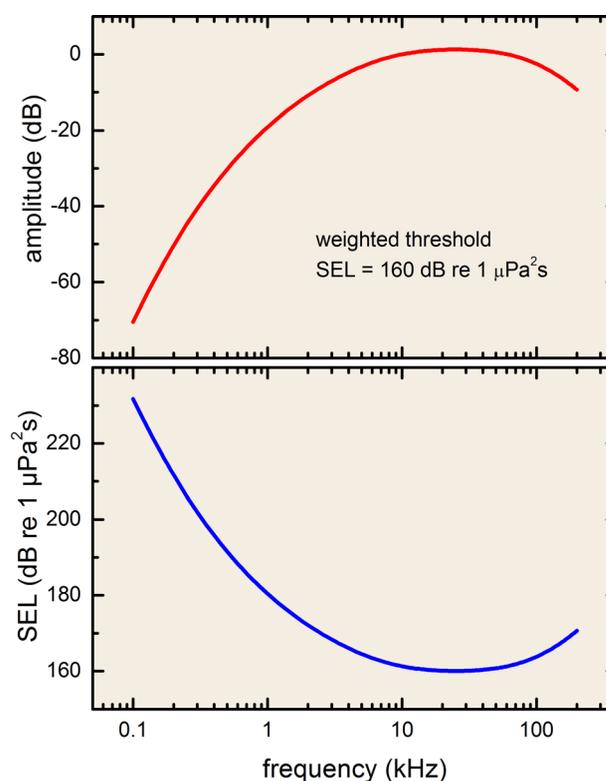


FIG. 2. (Color online) Hypothetical curves showing the relationship between (upper) an auditory weighting function and weighted threshold of NIHL and (lower) the resulting NIHL exposure function. For this example, unweighted noise at frequencies and levels above the exposure function would be predicted to result in NIHL.

provide a measure that reflects how sound affects hearing or behavior.

## II. WEIGHTING SOUND: HUMAN OVERVIEW

The development of several crucial sound metrics preceded the development of the type of auditory weighting function operation described in Fig. 1, when applied to human hearing. These include the development of the decibel as a measure of sound level, the development of a reference level when decibels are used to describe human hearing, the measure and standardization of the audiogram, the development of equal-loudness contours, and the development of frequency weighting functions based on equal-loudness contours.

Sound frequency and its perceptual correlate, pitch,<sup>1</sup> have been measured for many hundreds of years. The same is not true for sound level as it relates to amplitude, intensity, energy, power, or pressure (“sound level” here is not to be confused with the more specific term as defined in ANSI S1.1-2013; see the Appendix). It was not possible to accurately measure and control the level of sound until sound was generated by electronic means in the beginning of the 20th century;<sup>2</sup> well-controlled audiometric studies of level (e.g., audibility, level discrimination, loudness, and masking) did not occur until the 1920s. The first detailed measures of sound level, and hence the beginning of the use of weighting functions, started at the Bell Laboratories at this time (Gertner, 2012).

Figure 3 provides a timeline of events related to the development of auditory weighting functions for humans. Almost all of the initial research related to auditory weighting functions was performed at the Bell Laboratories from the early 1920s to the 1950s (Gertner, 2012; Yost, 2015). Many articles published by Bell Lab scientists during this time, as well as the notes from Bell Lab reports (Allen, 1996; Rankovic and Allen, 2001), indicate that key decisions were made at Bell Labs as to the best way to weight sound in order to improve telephone communication. As is described in the following, this work was highly relevant for human auditory perception and performance beyond telephone communications. Most of the weighting functions that evolved from the original Bell Lab work are now applied in very wide contexts, and sometimes without consideration of their appropriateness.

### A. Behavioral and psychophysical measures of sound perception in humans

While the sound weighting functions described later in this report are manipulations of the physical measures of

sound, the history of sound weighting functions and their validity for describing auditory processes is primarily based on human behavioral (often psychophysical) measures. To a lesser extent, physiological techniques can and have been used. There is a limited literature proposing specific weighting functions for non-human terrestrial mammals, such as that based on the rat audiogram (Björk *et al.*, 2000; Voipio *et al.*, 2006), but these will not be discussed in detail in this review. While not considered with respect to weighting functions, Fay’s (1988) classic vertebrate psychophysics data book contains behavioral data for a wide variety of stimuli and vertebrate species that might be relevant to developing weighting functions in non-human animals.

The history of obtaining behavioral measures of perception in a systematic way starts with the two-volume book by Fechner (1860), *Elemente der Psychophysik*. A century later, a paradigm shift in developing psychophysical methods developed with the introduction of the Theory of Signal Detection (TSD; Green and Swets, 1988). In between Fechner’s work and the emergence of TSD, a major set of psychophysical procedures (e.g., scaling procedures) were described by Stevens (1946).

In all of these procedures, a psychometric function is the basis of describing the relationship between physical measures of sound and measures of human perception or performance that result from the presentation of the sound. A psychometric function relates a measure of perception (e.g., loudness) or performance (e.g., percent correct discrimination) to a physical parameter of sound (e.g., sound level for loudness or a difference in sound level for percent correct discrimination). Weighting functions are based on a perceptual or performance value (e.g., loudness level) obtained from the psychometric function for a particular value of a sound parameter (e.g., sound pressure). In human hearing research the main dependent variable of the psychometric function is either a measure of a subjective aspect of sound [e.g., loudness in sones (see below)] or a measure of discrimination performance [e.g.,  $d'$  or percent correct responses]; however, other measures have occasionally been used. Regardless of the specific measure used, psychometric functions are always either explicitly or implicitly the basis of all auditory weighting functions.

From the time of *Elemente der Psychophysik* to today it has been recognized that measuring human perception of sound using behavioral procedures involves two classes of

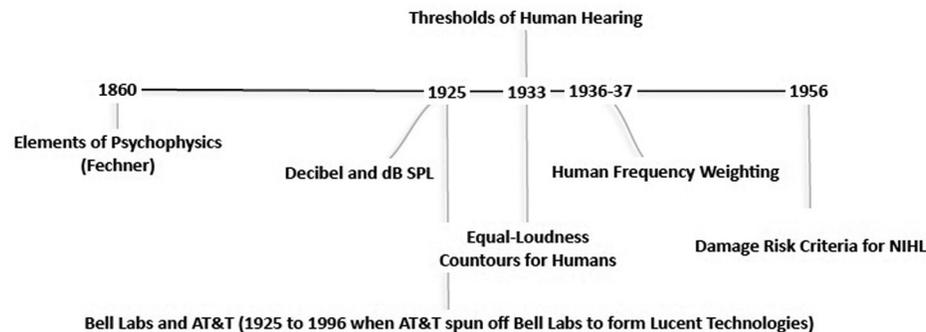


FIG. 3. Timeline of key reports relevant to the development of weighting functions for humans.

variables: the sensitivity of the human listener to one specific aspect of the sound that is presented, and the response bias or proclivity of the listener to use the various responses made available to him or her. Since weighting functions are intended to reflect aspects of sensitivity, influences of response bias on the obtained measures represent a confounding influence in establishing a psychometric function that only represents the sensitivity of the listener to the sound presented. The methods developed by Fechner and Stevens are biased measures of human auditory perception and performance, as the results that they produce can vary as a function of both the listener's sensitivity to sound and the listener's bias to various response options. On the other hand, TSD provides methods of bias-free estimates of sensitivity (e.g.,  $d'$ ). However, most bias-free measures assume that the underlying decision process employed by a listener is normally distributed—an assumption that is not always met.

Bias-free measures of human auditory perception or performance require a relatively large number of data points, which can take considerable time to obtain. Bias-dependent measures can be partially corrected to reduce the influence of biases and usually require far fewer data points. Most of the weighting functions described in this report were derived from bias-dependent measures, but often with bias-free measures collected at different times validating the weighting functions. Differences in weighting functions of human auditory perception or performance obtained between bias-free and bias-dependent measures are usually small. Thus, although bias-free measures should be used, when possible, efficiency often requires that bias-dependent measures be used. As an aside, the increased time requirement of bias-free measures—along with the challenges of training animals for psychophysical tasks—is the main reason why bias-dependent measures are much more common in marine mammal studies (see below).

## B. The decibel (dB)

Auditory weighting functions are based on the *decibel*. Compressive nonlinear transforms of sound intensity or pressure, expressed in units of dB, were first proposed at the Bell Laboratories. The precursor to the dB was called the transmission unit, or TU, which was itself preceded by miles of standard cable (MSC), where 1 MSC was the loss in power over 1 mile of length of standard telephone cable at a frequency of 795.8 Hz [see [Herbert and Proctor \(1934\)](#)]. It is crucial to recognize, and is evident in the evolution of sound level metrics, that the work at Bell Labs involving sound was focused almost entirely on telephone communication, and not on general principles of hearing. The TU was renamed/redefined in 1925 in honor of the laboratory's founder Alexander Graham Bell.

The decibel is a unit of *level*—the base ten logarithm of the ratio of a variable quantity to a corresponding reference value of sound power, or other power-like quantity, is a *bel* (B). In formulaic form, the level of a power-like quantity is defined, in bels, as

$$L = \log_{10} \left( \frac{Q}{Q_0} \right) \text{B}, \quad (1)$$

where  $Q$  is the power-like quantity and  $Q_0$  is the corresponding reference, or in decibels, as

$$L = 10 \log_{10} \left( \frac{Q}{Q_0} \right) \text{dB}. \quad (2)$$

The decibel is a relative unit, not an absolute unit. It is based on the ratio of two power-like quantities (e.g., sound power or intensity). That is, a sound with intensity of  $x$  dB is  $x$  dB more ( $x > 0$ ) intense, less ( $x < 0$ ) intense, or equal ( $x = 0$ ) in intensity to the reference sound. Thus, in expressing level in dB it is necessary to provide the reference. By adopting an appropriate reference, the decibel becomes a more meaningful number.

## C. Sound pressure level

In acoustics, the most commonly encountered level is *sound pressure level* (SPL), mathematically defined as (ANSI S1.1-2013)

$$SPL = 10 \log_{10} \left( \frac{\overline{p^2}}{p_0^2} \right) \text{dB}, \quad (3)$$

where  $\overline{p^2}$  is the time mean-square sound pressure and  $p_0$  is the reference pressure. Mainly because a change in SPL of one bel produced a large perceptual difference, it was decided that the decibel would be a more appropriate and useful measure of SPL for telephone communication. Another important and related measure is the *peak sound pressure level* (pSPL), here defined as

$$pSPL = 10 \log_{10} \left( \frac{\max p^2}{p_0^2} \right) \text{dB}, \quad (4)$$

where  $p^2$  is the squared instantaneous sound pressure and  $p_0$  is the reference pressure.

For SPL, the reference pressure is 20  $\mu\text{Pa}$  for measurements in air or gases, and 1  $\mu\text{Pa}$  for underwater measurements in the International System of Units (SI)<sup>3</sup> (ANSI S1.1-2013). The currently recommended reference values have, however, gone through several iterations. In the original work done at Bell Labs to measure human sensitivity to sound ([Sivian and White, 1933](#)), the reference pressure was 1  $\mu\text{bar}$ . As a result, the thresholds of human hearing were near  $-80$  dB re 1  $\mu\text{bar}$ . To make the threshold of hearing closer to 0 dB, rather than  $-80$  dB re 1  $\mu\text{bar}$ , for a healthy, normal-hearing young person in the frequency range of best hearing, a reference pressure of 20  $\mu\text{Pa}$  was adopted (see below). This reference was used in the first sound level meter standard, Z24.3-1936 ([Acoustical Society of America, 1936](#)).

While the main rationale for using measures like the transmission unit, bel, or decibel is the reduction of a large range of intensities to which humans are sensitive (i.e.,  $10^{13}$ )

to a more manageable range (i.e., 130 dB), the decibel scale offers two other related advantages in terms of representing human sensitivity to sound. First, the human auditory system is sensitive to proportional changes in intensity or pressure, and the log scale used for dB measurements is a proportional scale. Second, it is fortuitous that a 1-dB change in SPL is approximately a just-noticeable difference for humans across a wide range of levels and frequencies (e.g., [Jesteadt et al., 1977](#)). Thus, the decibel scale reflects the human auditory system's sensitivity to both level and changes in level.

The use of SPL in units of dB was reinforced by the collection of hearing measurements from approximately 400 000 people who attended the 1939 World's Fair in New York and San Francisco and participated in a brief hearing test administered by employees of the Bell Labs, then called the Bell System [see [Steinberg et al. \(1940a\)](#)]. The average threshold of hearing for people in the range of 10–19 years of age (7549 people) was 18 dB re 20  $\mu$ Pa at 1760 Hz (the frequency with the lowest mean threshold of the five frequencies tested). The threshold of hearing increased as the age of the participants increased and when the test frequency was lower or higher than 1760 Hz ([Steinberg et al., 1940b](#)). Given the conditions under which the World's Fair hearing tests were conducted, the proximity of the thresholds to 0 dB re 20  $\mu$ Pa indicated that 20  $\mu$ Pa was a reasonable reference sound pressure for indicating the minimum sound pressure the average young human with normal hearing could detect. Thus, expressing sound magnitude in terms of SPL, with units of dB using a 20- $\mu$ Pa reference, represented a step beyond the decibel scale in weighting measures to reflect human sensitivity to sound. Considerable effort was given to developing the decibel as a unit of sound magnitude for the purpose of characterizing human hearing. It is not always clear that the decibel is the best unit to use for all other animals, but the decibel has been essentially universally adopted anyway.

#### D. Sound exposure level

A level quantity often used for auditory damage risk criteria, especially for marine mammals, is the *sound exposure level* (SEL). This metric takes both level and duration of sound into account. Sound exposure is the time integral of the square of the instantaneous sound pressure. SEL is ten times the logarithm to the base 10 of the ratio of a sound exposure to the reference value

$$SEL = 10 \log_{10} \left( \frac{\int_0^T p^2(t) dt}{p_0^2 T_0} \right) \text{dB}, \quad (5)$$

where  $p(t)$  is the instantaneous sound pressure,  $t$  is time,  $T$  is the duration of a stated time interval,  $p_0$  is the reference value for sound pressure, and  $T_0 = 1$  s is the reference duration for sound exposure level (ANSI S1.1-2013). (As noted, SEL has had greatest relevance to the development of weighting functions in marine mammals and receives greater attention in the section on marine mammals.)

#### E. Standardized thresholds of hearing and dB hearing level (HL)

Threshold of hearing represents the primary description of the sensitivity to sounds of different frequencies, which is the base relationship behind most auditory weighting functions. That is, threshold of hearing is frequency dependent but the frequency dependence changes as SPLs increase above threshold. Measures of equal loudness phenomena reflect these frequency-specific dependencies. In humans, the threshold of hearing is often measured at octave frequencies ranging from 125 Hz to 8 kHz, and usually includes selected (near) half-octave frequencies such as 750, 1500, 3000, and 6000 Hz. Starting with [Sivian and White \(1933\)](#),<sup>4</sup> hearing threshold was either measured using a loudspeaker in an open field [now called a minimal audible field (MAF) measure], or using headphones (or insert ear buds) and typically quantified by measures in an acoustic coupler [now called minimal audible pressure (MAP) measures]. MAF and MAP thresholds have been standardized for a range of frequencies both nationally in the U.S. (ANSI S3.4-2007) and internationally (e.g., ISO 389-8:2004). That is, there has been a consensus agreement within a standards process that the average otologically normal-hearing person's thresholds of hearing (expressed in dB SPL) are specific values. MAF measures are the ones most relevant to the focus of this review.<sup>5</sup>

The standardized method for estimating MAF threshold requires that the distance from the listener to the loudspeaker be less than 1 m and that the acoustic field is free from any major reflective surfaces. The duration and rise-fall time of the tone, along with the psychophysical procedure used to obtain thresholds, are specified in the associated standards. Using standardized approaches is a necessity for comparability across studies and in establishing baseline and reference values for comparisons among individuals.

Hearing level (HL) is the amount by which the hearing threshold for a listener exceeds a specified reference equivalent threshold level. The threshold is usually expressed in dB SPL and measured via standardized MAF techniques. Because the standardized threshold of hearing changes with frequency, dB HL is a measure of sound level that is weighted based on frequency and the mean or median threshold of hearing for young, otologically normal listeners. HL is the measure used to define hearing loss for clinical (audiological and otological) situations and for most research investigations. Audiometers, which are the instruments used to measure thresholds of hearing in a clinical setting, are calibrated based on a particular standard and its standardized thresholds of human hearing (e.g., ANSI S3.6-2010). Note that the normative thresholds that dB HL is based on vary not only with frequency, but also with the ear-phone and coupler used during testing. If a patient has a threshold of hearing of 60 dB HL at 4 kHz, their threshold of hearing at 4 kHz is 60 dB greater than that of the normative threshold of hearing at 4 kHz. Thus, dB HL > 0 dB represents a hearing loss, dB HL = 0 dB represents a threshold of hearing that is ideally normal, and dB HL < 0 dB represents a threshold of hearing that is lower than that specified in the

standard. From a clinical perspective, the dB HL values that represent relatively “normal” hearing are typically considered to be thresholds of 20–25 dB HL or lower.

## F. Equal loudness contours and the phon and sone scales of loudness

The audiogram describes a type of auditory weighting function for sounds that are (by definition) just barely audible. The sounds that might adversely affect an animal are generally more intense or judged to be subjectively loud. Fletcher<sup>6</sup> found that the perceived loudness of tones as a function of changing physical sound level seemed to depend on the tone’s frequency. Fletcher and Munson (1933) subsequently measured perceived loudness as a function of tonal level and frequency. Listeners determined the level of a tone of a particular frequency (a comparison tone) that they judged equal in perceived loudness to a 1000-Hz standard tone presented at a particular level and expressed in dB SPL. Each point on any particular equal loudness contour represented the levels (y-axis) and frequency (x-axis) that were judged by listeners to be equally as loud as the 1000-Hz standard at the specified SPL (see Fig. 4). As a consequence, all tones with the levels and frequencies indicated by a particular equal loudness contour were considered equally loud to each other. Note that as Fletcher informally observed, the loudness of low-frequency comparison tones changed more rapidly with changes in the level of the standard tone than did tones in the mid-frequency range near the 1000-Hz standard. Changes in loudness for high-frequency comparison tones also changed more rapidly with changes in the intensity than that which occurred for mid-frequency tones, but less so than for low-frequency comparison tones. A logical

conclusion of these findings was that the physical level of a sound was not a good measure of perceptual loudness, since loudness changed as a function of frequency and physical level did not.

The “phon,” as a measure of “loudness level,” was defined based on the equal-loudness contours (Fletcher and Munson, 1933). A sound with a loudness of “n” phons is a sound that is judged equally loud to a 1000-Hz tone presented at “n” dB SPL. Stevens (1946) later measured loudness using the “sone” scale. A sound of “n” sones is “n” times louder than a sound with a loudness level of 40 phons, which is defined as having a loudness of 1.0 sone (e.g., a sound of 3 sones is judged 3 times louder than a 1000-Hz tone presented at an SPL of 40 dB re 20 μPa). The relationship between sones and physical sound level is often expressed as a power function [Stevens power law, see Stevens (1946)] when the exponential of the power function is approximately 0.3 (i.e., a 10-dB change in physical sound level is approximately a doubling of perceived loudness in sones). The power function fit of sones versus phons is approximately linear for loudness levels above about 40 phons when sones are plotted on a logarithmic scale. The slope of the linear fit is about 0.3, and as a result, sones (S) can be related to phons (P) by the following formula:

$$S = (10^{(P-40)/10})^{0.3} \approx 2^{(P-40)/10}. \quad (6)$$

Studies of equal-loudness measures like Fletcher’s and Munson’s have been criticized for several reasons (see below). As a result, Robinson and Dadson (1956) obtained a new set of equal-loudness curves for people tested with a frontal sound source in an anechoic chamber. The Robinson-Dadson curves were standardized as ISO 226 in 1986. In 2003, ISO 226 was revised again based on equal-loudness contours collected from 12 international studies (see Fig. 4, which shows the differences between the Robinson-Dadson curve and equal-loudness curves from ISO 226:2003).

Since the development of equal-loudness contours and the measurement of loudness for tones, standards for calculating the loudness of stationary noise sources have also been developed. Over the years there have been three procedures that have been standardized for measuring the loudness of noise: The first method, in ISO 532, method A, was based on the work of Stevens, while the second method, method B, was based on the work of Zwicker [and see also Deutsches Institut für Normung eV (DIN) 45631]; and since then (the third method), ANSI S3.4-2007 included the loudness model of Glasberg and Moore (2006). All of these loudness measurement approaches use data on equal-loudness contours obtained for 1/3- or 1-octave wide white (or pink) noises with different center frequencies to develop a method for calculating the loudness of stationary noise sources. These approaches are equivalent to standardized weighting functions and are used to estimate human loudness perception in many different contexts. The relationship of equal loudness contours to sound weighting functions is explained in Sec. II G.

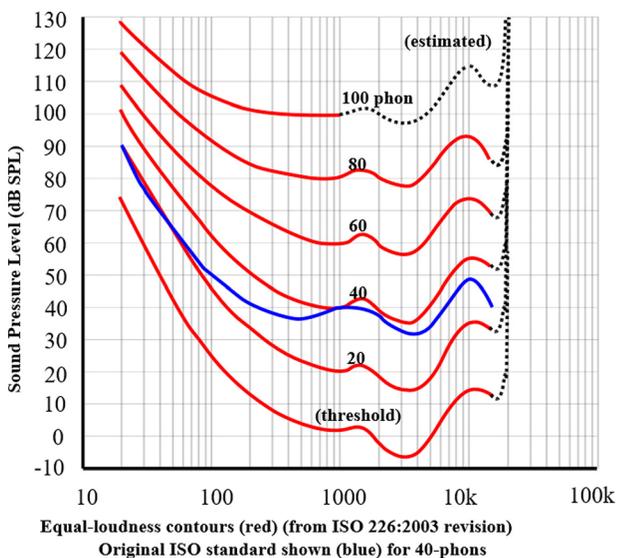


FIG. 4. (Color online) Equal loudness contours from human listeners. Equal loudness contours from Robinson and Dadson (1956) are plotted in blue, and the equal loudness contours from the ISO 226:2003 standard are plotted in red. Equal loudness contours are shown for the 0 (threshold), 20, 40, 60, 80, and 100 phon levels. This figure was created by Peter J. Skirrow and is distributed under the Creative Commons Attribution-Share Alike 3.0 Unported License.

## G. Auditory weighting functions

### 1. A-, B-, C-, D-, Z- weighting functions

Approximately three years after the publication of the paper of Fletcher and Munson (1933) on tonal equal loudness contours, a sound level meter standard (Z24.3-1936) was developed under the auspices of the American Standards Association (later renamed the American National Standards Institute) that included a frequency weighting of sound (Acoustical Society of America, 1936). That is, a form of filtering could be imposed before a sound level reading was made such that frequencies to which the auditory system was insensitive (or less sensitive) would be filtered (i.e., their contributions reduced) from the overall sound level measurement (see Fig. 1). Over the next few decades [and for subsequent revisions of that standard, which became ANSI S1.4 and then International Electrotechnical Commission (IEC) 61672-1] both national and international standards added several different weighting schemes (filters) for sound level meters. These weighting schemes included A-, B-, C-, D-weighting, which are largely based on equal-loudness contours, as well as Z-weighting, which has a flat frequency response. Today, the B and D weightings are no longer included in the standards, largely because of the broad use of A-weighting (discussed later).

In simple terms, a weighting curve is a bandpass filter that is an inverted and smoothed equal-loudness contour (see Fig. 1). The inverse of the A-weighting curve, which is the most used auditory weighting function for a wide variety of human activities involving sound exposure, compares well with the 40-phon equal-loudness contour of Fletcher and Munson (1933). Figure 5 shows the A-, B-, C-, and D-weighting curves as filter functions. The C-weighting curve has a shallower low-frequency roll-off than the A-weighting curve because the C-weighting curve is derived from the 100-phon equal loudness contour, which is flatter (i.e., sound

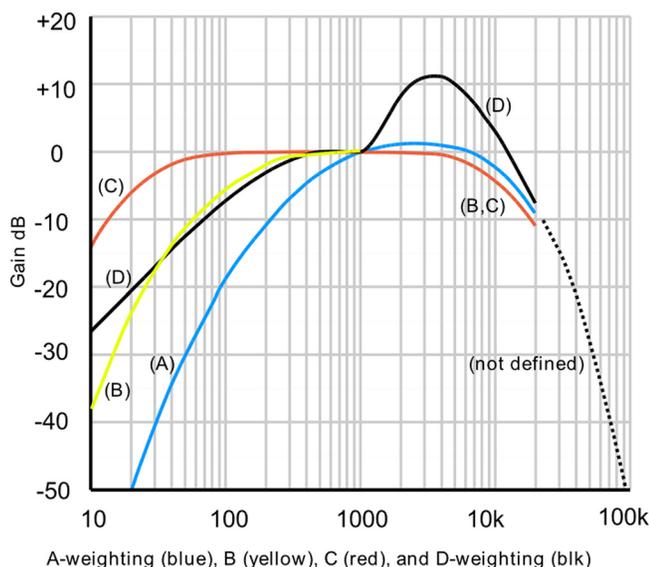


FIG. 5. (Color online) A- (blue), B- (yellow), C- (red), and D- (black) weighting functions are shown. This figure was created by Peter J. Skirrow and is distributed under the Creative Commons Attribution-Share Alike 3.0.

level does not change as much with frequency for 100-phon sounds as it does for 40-phon sounds). Thus, the C-weighting function more heavily “weights” low-frequency sounds than A-weighting and represents a better estimate of the auditory system’s response to high-level sounds than does A-weighting (at least in terms of subjective magnitude). A-weighting was originally designed for measurements at low sound levels (e.g., 40 phon sounds), but it has been applied in many procedures and policies as a measure for almost any sound level. The Z-weighting function, also known as zero weighting or the linear scale, is a broadband filter with a flat frequency response (within  $\pm 1.5$  dB) from 10 Hz to 20 kHz, that can be used for measuring the level of transient stimuli or the overall level of a broadband stimulus, such as speech or masking noise.

### 2. Other weighting functions

While functions based on equal loudness contours are the most common form of auditory weighting functions, several other “forms” have been developed for humans (and other animals). For example, note that the 0-phon curve (zero loudness curve) is very much like the threshold of hearing curve [Fig. 4; see Buus *et al.* (1998) for a discussion of loudness at threshold], although it is anchored to the threshold of hearing at 1000 Hz, rather than 0 dB SPL. Thus, it is possible that a spectral weighting function could be based on the threshold of hearing, although this would be most appropriate for extremely soft sounds. No such weighting function has been developed for humans, although Björk *et al.* (2000) and Voipio *et al.* (2006) refer to such a human threshold weighting function as dBH (also see Sec. V A for a discussion of “dBht” weighting).

The International Telecommunication Union-Radio Communication Sector (ITU-R) 468 weighting standard is widely used in measuring noise and the recording of sounds for music and other purposes. The ITU-R 468 weighting function factors in aspects of noise that A-weighting does not, since the A-weighting curve was based on experiments using only tones. Figure 6 compares the ITU-R 468 weighting function, the ISO 226-2003 40-phon equal loudness contour (inverted), and the A-weighting function.

### 3. Other psychoacoustic measures that are similar to loudness when sound level and frequency vary

A main characteristic of equal-loudness contours is that loudness is frequency dependent when overall sound level is low, but less frequency dependent when overall sound level is high (i.e., equal loudness contours become more “flat” as a function of frequency as overall sound level increases). Similar effects have been measured for many psychoacoustic phenomena including intensity discrimination (Riesz, 1928; Jesteadt *et al.*, 1977), frequency discrimination (Shower and Biddulph, 1931; Wier *et al.*, 1977), and masking (Reed and Bilger, 1973). Intensity discrimination thresholds can be considered as an example. In both the original study of intensity discrimination (Riesz, 1928) and in a more recent comprehensive study (Jesteadt *et al.*, 1977), the thresholds for discriminating a change in sound level in decibels (i.e., the

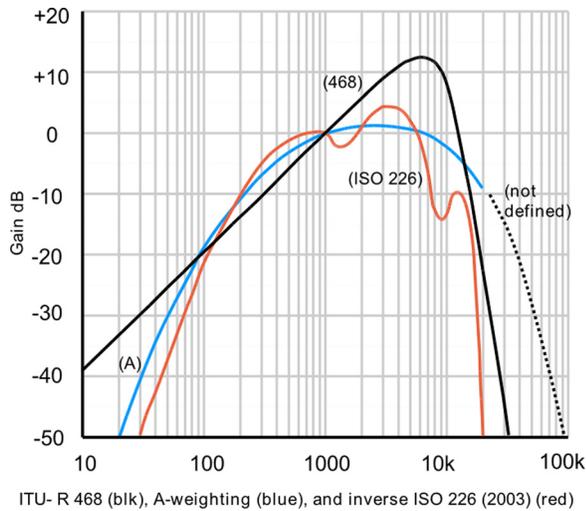


FIG. 6. (Color online) The A-weighting function (blue), the inverse of the 40-phon ISO 226 (2003) Equal Loudness Contour (red), and the ITU-R 468 filter function are shown. This figure was created by Peter J. Skirrow and is distributed under the Creative Commons Attribution-Share Alike 3.0 Unported License.

threshold difference in the level of two successively presented tones), varied both as a function of tonal frequency and tonal sound level. In the *Jesteadt et al. (1977)* study, the mean level difference required for a level discrimination at low overall sound levels (SPL of 40 dB re 20  $\mu$ Pa) was 1.8 dB for an 8000-Hz tone, 1.1 dB for a 1000-Hz tone, and 1.4 dB for a 400-Hz tone. At a higher overall level (SPL of 80 dB re 20  $\mu$ Pa), tones at all frequencies required about a mean 0.9 dB difference in sound level for discrimination. Thus, as is the case for loudness, intensity discrimination threshold is frequency dependent at low overall levels, but not at high overall levels. Many of these level-frequency interactions can be explained by peripheral sound processing [see *Yost et al. (1993)*] in ways that are similar to the model of loudness of *Glasberg and Moore (2006)* (see below). One conclusion of this observation is that auditory weighting functions might be derivable using psychoacoustic phenomena other than loudness.

#### 4. Measuring loudness for human subjects using behavioral methods

Since A-weighting—based on the 40-phon equal-loudness contour—is so widely used for sound measurement, it is important to understand how loudness is measured for humans. Loudness is a perceptual/subjective attribute of sound magnitude, and is defined in ANSI S1.1-2013 (definition 11.03) as, “the attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud.” In contrast, sound intensity, sound pressure, sound energy, and sound power are physical/acoustical measures of sound magnitude. Thus, loudness can be used to indicate how auditory perception is influenced by the physical/acoustical properties of sound. Knowing just the physical/acoustical properties does not allow for accurate accounts of auditory perception, because the physical/acoustical properties of

sound interact with a receiver’s auditory system and perceptual experience.

Loudness has been measured psychophysically in humans using (mostly) one of three psychophysical procedures:

- (1) **Matching:** Listeners match two sounds in terms of their perceived loudness (for example, how intense does sound A have to be to be perceived as equal in loudness to sound B?).
- (2) **Scaling:** Listeners rate/scale the perceived loudness of a sound (for example, on a scale of 1 to 100, where 1 is just audible and 100 is uncomfortable, how loud is sound A?).
- (3) **Comparisons:** Listeners indicate which of two or more sounds is louder/softer (for example, is sound A or B louder?).

*a. Matching.* Equal-loudness contours are the primary data resulting from using a matching procedure. Equal-loudness contours demonstrate that loudness is both level and frequency dependent, as opposed to physical/acoustical measures of level and frequency, which are independent. For instance, by definition, the subjective/perceptual loudness level of a 1000-Hz tone varies from 20 to 100 phons as the physical level of the tone varies from 20 to 100 dB SPL (re 20  $\mu$ Pa; 80-dB range). The subjective/perceptual loudness of a 100-Hz tone varies from 20 to 100 phons as the physical level of the tone varies from 40 to 105 dB SPL (re 20  $\mu$ Pa; 65-dB range).

*b. Scaling.* Various scaling procedures have been used to measure loudness. These procedures are response-based or stimulus-based. In a response-based scaling procedure, listeners provide an estimate using a number, a descriptor (e.g., loud or soft), or some other stimulus value (e.g., the brightness of a light) to indicate the perceived loudness of a sound. In stimulus-based scaling procedures, listeners adjust the level of sound (or some other physical/acoustical parameter) to indicate perceived loudness. For example, a listener may increase the level of a sound by 10 dB to make it twice as loud as a standard sound.

*c. Comparison.* A wide variety of procedures ask listeners to determine if one sound is louder or softer than another sound. Collectively, the major result of such comparisons is that sound loudness varies as a function of several different physical/acoustical and contextual variables, including duration, bandwidth, and temporal envelope of the sound, the context of the presentation, and whether the presentation is monaural or binaural. These variables are explained more below.

- (1) **Duration:** In humans, short sounds are perceived as softer in loudness than long sounds for durations less than approximately 500 ms. Changes in loudness with duration are due to temporal integration. For sounds longer than approximately 500 ms, sound loudness does not appear to increase as duration is increased. The change in loudness approximates an equal-energy rule in that

sounds of the same energy are often judged to be equal in loudness up to about 500 ms in duration. For longer durations, sounds with the same power tend to be judged equal in loudness. A temporal integrator can be used to model the relationship between perceived loudness and duration [at least for simple sounds like tones, Viemeister and Plack (1993)].

- (2) Bandwidth: Sound loudness is often determined by the sound's frequency bandwidth relative to the critical bandwidth for the particular sound (see below). In humans, sounds of similar frequencies are perceived as louder when they are combined than when either sound is presented individually, while sounds of disparate frequencies are not perceived as louder when combined as compared to individual presentations. Loudness summation occurs when the frequency difference between the sounds is less than a critical band (at the average frequency; see more on critical band below). There is no or limited loudness summation when the frequency separation is greater than a critical band (Florentine *et al.*, 2011).
- (3) Temporal envelope: Sounds with long rise times and fast decay times are often judged to be softer in loudness than sounds with sudden onsets and slow decay times, even when all other aspects of the sounds are the same (Florentine *et al.*, 2011).
- (4) Binaural presentation: The same sound presented to two ears is often judged to be louder than when the sound is presented to only one ear. The difference in loudness is usually less than the equivalent of a 3-dB change in loudness (Florentine *et al.*, 2011).
- (5) Context: Perceived loudness of sounds can depend on the context in which loudness is being judged. For example, the same sound presented in the context of soft sounds is often judged to be louder than when it is presented in the context of louder sounds (Florentine *et al.*, 2011).

## 5. Loudness models

There are three models of loudness that have been used to develop standard methods for measuring loudness: Stevens method, Zwicker method (ISO/CD 532-1), and Glasberg and Moore (GM) method. The GM model is the most appropriate one for establishing spectral weighting functions for dealing with human perception. The GM method argues that loudness based on changes in level and frequency is determined by processing information about sound in the outer, middle, and inner ear and relayed to the brain by the auditory nerve. That is, the neural information provided to the brain by auditory nerve fibers reflects sound loudness and not just the physical/acoustical sound level. Thus, in this model, the neural activity is dependent on both physical/acoustical level of the sound and sound frequency, not just on the physical/acoustical level of the sound.

Details of the GM model can be found in Glasberg and Moore (2006). The model uses bandpass filtering to simulate the change in level with frequency as sound passes through the outer and middle ears. The next stage in the model is a

simulation of cochlear tuning resulting in excitation patterns for the spectral components of a sound. Then there is a simulation of the nonlinear compressive transform of sound vibration to neural action potentials based on a combination of cochlear outer hair cell function and the generation of action potentials within the auditory nerve. These stages produce a specific loudness pattern for the sound. Loudness measured in sones is assumed to be the area under a specific loudness pattern. Thus, if the outer and middle ear transfer functions, cochlear tuning, and cochlear nonlinearity are known, the loudness of a sound in sones and equal loudness contours (phons) can be derived for sounds with simple spectra, especially tones and stationary white or pink noise.

## 6. Loudness procedures when human subjects, behavioral procedures, or direct estimates of loudness are not used

Based on the ANSI definition of loudness (Sec. II G 4), loudness cannot be determined unless a sound can be judged as being "soft" or "loud." As a result, loudness cannot technically be measured in non-human animals or via methods that do not allow for such judgments to be "stated" by the subject (e.g., as in physiological measures, such as evoked potentials). However, indirect measures of loudness have been obtained. The key aspects of indirect measures are the degree to which they are valid and reliable analogous measures to the subjective/perceptual behavioral measures obtained from human listeners. The validity and reliability for indirect measures are usually evaluated by comparing the resulting data to those obtained from a direct measure when human listeners are used. For non-human listeners, the step described above is performed with human listeners and then the indirect measure is applied with an animal subject and results compared to those obtained using the measure with human listeners.

The crucial aspect of the comparison is the degree to which the indirect measure varies in the same way as a function of physical/acoustic variations applied to the sound whose loudness is being estimated. For instance, the variation in loudness for human listeners as a function of level and frequency occurs in a certain way. Thus, for a proxy method to be useful, the change in the indirect measure of loudness should follow the same pattern that exists for a direct measure of loudness when frequency and level are varied. In particular:

- (1) Loudness varies more with tonal frequency with a fixed, low-intensity tone, than it does when the intensity of the tonal sound is high in level.
- (2) For a fixed sound intensity (especially at low intensities), low- and high-frequency tones are perceived as softer than mid-frequency tones. Note that as species have different hearing profiles, low-frequency sounds can be considered those in an animal's region of diminishing sensitivity, mid-frequency sounds are those near an animal's region of best sensitivity, and high-frequency sounds are those in the upper region of diminishing sensitivity. For example, in toothed whales, a mid-frequency sound would be well above the highest-

frequency sound a human can hear. So mid-frequency should be related to the region where a given animal species hears best, low-frequency and high-frequency regions where the sound is still audible, but at (substantially) higher thresholds than observed in the “mid-frequency” region.

Thus, to identify an appropriate indirect measure of loudness, the same physical sound variables should be used with the indirect loudness measure as were used with a direct measure and the results should generally follow the two relationships (1 and 2) stated above.

Loudness matching, scaling, or comparison cannot usually be used to obtain an indirect measure of loudness. Typically, in indirect measures, the level of a sound is varied for each of a variety of sound parameters (e.g., frequency or duration). The value of the indirect measure is obtained as a function of stimulus level for each value of the physical/acoustical variable (e.g., tonal frequency). Then, the stimulus level required for a specified value of the indirect measure is used to scale the indirect measure of loudness. As an example, reaction time (time between a sound presentation and a subject’s response) has been used as an indirect measure of loudness. In these experiments, reaction time is measured as a function of tonal sound level for each of a number of different tone frequencies. The level of the tone at each frequency required for particular reaction times is determined to establish a constant reaction-time contour; for example, what level was required at each frequency for a reaction time of 2 s, or 1 s, or 0.5 s, etc.? This relationship is then viewed as an indirect measure of loudness. While such an approach can be useful, the validity of the indirect measure as an indicator of loudness depends on the extent to which the constant reaction-time functions follow the two relationships stated above, i.e., (1) Does the level required for a constant reaction time vary more as a function of frequency when the overall physical/acoustical level is low as compared to when it is high? (2) Is the physical/acoustical level of tones required for a constant reaction time higher for low- and high-frequency tones as compared to mid-frequency tones?

Based on the results from experiments designed to obtain an indirect measure of loudness when frequency and level vary, one of three conclusions can be reached:

- (1) The data regarding the indirect measure are not consistent with loudness data from human listeners (e.g., the indirect measure depends on frequency in the same way at all physical/acoustical levels).
- (2) The data regarding the indirect measure are generally consistent with data from human listeners in that the indirect measure changes with both frequency and level (weak evidence for an indirect measure of loudness).
- (3) The data regarding the indirect measure are consistent with data from human listeners in that the indirect measure changes with both frequency and level, such that the indirect measure shows a greater dependency on frequency at low overall levels than at high overall levels, and low- and high-frequency sounds require higher levels than than mid-frequency sounds for equal indirect

measurement values (strong evidence for an indirect measure of loudness).

Before providing a more detailed review of the literature on reaction time and physiologic measures of loudness, a few additional caveats about indirect measures of loudness are worth reemphasizing. First, indirect measures are not true measures of loudness because they are not subjective measures of soft and loud. Second, for human listeners, a wide range of stimulus and contextual variables influence perceived loudness. The extent to which a particular indirect measure of loudness is consistent with human listeners’ perception of loudness will probably depend on the extent to which the indirect measures also vary as a function of these stimulus and contextual variables.

### 7. Reaction time as an indirect measure of loudness

The literature on reaction time and loudness primarily concerns how reaction times vary with frequency and/or level. Studies have included human listeners (including young children) and several animal models (e.g., monkey, cat, porpoise). Among the early studies, that by Chocholle (1940) is most often cited. Chocholle (1940) concluded that reaction times could be a good indirect measure of loudness in human listeners, but Kohfeld and colleagues later noted that there was a possible confound in Chocholle’s experiments in that a click was present in the presentation of several of his stimuli (Kohfeld *et al.*, 1981). They did not obtain the same results as Chocholle (1940) with click-free sounds leading to the conclusion that reaction time may not be a good indirect measure of loudness at low sound levels.

Stebbins (1966) varied frequency from 0.25 to 15 kHz and varied SPL over a 70 dB range while measuring the reaction time in two male *Macaca irus* monkeys. Reaction time decreased with increasing tone level. Equal latency contours were created for each of the two monkeys tested. In one monkey, the equal latency functions more or less paralleled one another, while in the other, the functions were more or less parallel at low and mid-frequencies, converging at higher frequencies. Thus, in one monkey, the parallel functions do not reflect human equal-loudness contours, and in the other monkey, convergence of functions in the low frequencies would be predicted from human loudness contours. Hence, the reaction time data in monkeys by Stebbins (1966) do not appear to resemble closely the equal loudness contours observed in human subjects. May *et al.* (2009) used the domestic cat (*Felis catus*) to assess reaction time. After baseline reaction times were collected, four cats were exposed to a 50-Hz bandwidth noise centered at 2 kHz and with a SPL of 109 dB re 20  $\mu$ Pa for 4 h. Following the noise exposure, threshold elevation from 2 to 4 kHz was  $\sim$ 20 dB. Equal latency contours from their pre-noise-exposure data appeared similar in shape to the equal loudness contours reported in humans. Following noise exposure, there was a steeper decrease in reaction time with increasing tone level, a phenomenon reminiscent of loudness recruitment that is often reported in humans with sensorineural hearing loss. Similar work has been conducted in marine mammals

(without a noise exposure component), and the results of these studies are reported in more detail later in this review.

Luce (1986) discussed in some detail the assumptions about what changes in reaction times might represent in terms of “mental” processing. Reaction times have at least three components: (1) the time it takes for the stimulus to be processed at an initial receptor stage and for this neural information to be sent to some decision processor; (2) the time it takes for the decision to occur; and (3) the time it takes to produce a motor output indicating the decision that was reached. It is often assumed that the first and last times are fixed for any one stimulus (e.g., a sound)/response (e.g., a button push) condition and as a consequence changes in reaction time for a fixed stimulus/response combination reflect the time it takes to make the appropriate decision. That is, changes in the reaction time to sound might represent the time to make a sound-detection decision. However, the transduction time of sensory information is inversely proportional to stimulus intensity. This is seen, e.g., in the decrease in latency with increasing stimulus level for the auditory brainstem response (ABR). For human ABRs, this change in latency is on the order of several ms for clicks over a 50 dB range (Burkard and Don, 2015). Thus, changes in reaction time when sound level is varied could be based on a combination of sensory transduction time and signal-detection decision time. Models of loudness (see above), especially the Glasberg and Moore (GM) model (2006), indicate that the information in the auditory nerve regarding sound magnitude depends on sound loudness. If time of transmission of the first component of reaction times is proportional to the neural indication of sound magnitude in the auditory nerve, then reaction time could reflect loudness, but only if the two additional stages (decision and motor output) are independent (e.g., fixed) of stimulus level and frequency.

Several studies of reaction time in human and animal models have suggested that changes in reaction time might reflect changes in loudness. Reaction time in many of these studies does depend on both tonal frequency and level, and there is some evidence that the change in reaction time with frequency is greater at low levels than at high sound levels, thus lending some support to the validity of reaction times as an indirect measure of loudness. However, reaction time is highly variable within and across listeners, much more so than direct measures of equal loudness contours, which are also prone to variability. Such variability makes it almost impossible to derive equal loudness contours that are reliable across a population of subjects. The best outcome of such cross-sectional studies is that reliable equal reaction time contours are obtained for an individual subject and, if so, serve as an indirect measure of an equal loudness contour for that subject. This may not reflect other subjects and is unlikely to provide a reliable indirect estimate of loudness relevant to a population of subjects.

The decision process for detecting sound is usually assumed to consist of two independent components [see Green and Swets (1988)]: a sensitivity component and a response-bias component. TSD usually assumes that sound loudness relates to sensitivity, and other variables (e.g., stimulus expectancy) relate to response bias. If appropriate

measures of detection are employed (e.g.,  $d'$ ), loud sounds lead to higher levels of detection performance than soft sounds, independent of response bias. It is possible that the decision process is faster the more sensitive the decision process is to the sound, although loudness is not a prediction of TSD. If so, then reaction time could depend on both the sensory processing of stimulus level as described by a loudness model and decision processing time.

But how does reaction time interact with response bias? Quoting from Emmerich *et al.* (1972), page 65, “Gescheider *et al.* (1969) found, in an experiment on the detectability of vibrotactile signals, that responses with lower probabilities of occurrence tended to have longer latencies. Response probabilities were manipulated by changing the *a priori* probability of signal presentation. It was predicted that when the inputs leading to a given response were close to the criterion for that response, along some decision axis, the latencies for that response would be unusually long. The data supported this prediction, showing long latencies, for example, for ‘yes’ responses under a strict decision criterion and for ‘no’ responses under a lax decision criterion.” In other words, reaction times seem to indicate, at least in part, a listener’s response bias. Emmerich *et al.* (1972) and others have derived latency receiver operating curves for detecting tonal sounds in noise for human listeners showing that changes in response times can reflect changes in response bias, at least to a first approximation. While the literature suggests that changes in response times are not exactly the same as changes in other ways of measuring response bias, this literature does pose a challenge for using reaction times as a measure of loudness; specifically, does a change in reaction time indicate that there is a change in loudness, a change in response bias, or both? If one can hold response bias constant, then reaction time might be able to indicate loudness as indicated above. It is not always clear that this can be done, especially in a non-human animal model.

## 8. Physiological measures of loudness

The use of auditory evoked potentials (AEPs) as a possible surrogate measure of loudness has the advantage of not requiring a motor or cognitive response. Thus, the cooperation of the subject is not needed and the time required for a motor response is not a confounding factor in estimating loudness, as is the case for reaction time. Although there are different AEPs that can be measured, short-latency responses (such as the ABR) offer the advantage of being strongly dependent on the physical aspects of the stimulus (including stimulus frequency and stimulus level). There are also minimal effects of sedation/anesthesia on the ABR (assuming normothermia is maintained), and hence, the short-latency AEPs offer advantages for use in subjects where sedation/anesthesia is desirable (or required). Longer-latency responses (such as cortical auditory evoked potentials) are much more dependent on subject state (anesthesia, attention, arousal) and are perhaps less desirable to use as surrogate measures of loudness. For human AEPs, response peak latencies decrease and amplitudes increase with increasing stimulus level, and hence, either the latency or amplitude of

specific waves might be used as a surrogate measure of loudness.

Several studies have compared loudness functions and ABR peak latency and/or amplitude changes across click level in humans. In some of these studies, the exponents of perceptual loudness were more closely approximated by the ABR peak amplitude data than the ABR peak latency data (Pratt and Sohmer, 1977; Wilson and Stelmack, 1982). In a reanalysis of the Pratt and Sohmer (1977) data, it was reported that the latency-intensity power function coefficients closely approximated the loudness growth coefficients (Babkoff *et al.*, 1984). Soon afterwards, it was found that in normal hearing subjects, the click-evoked wave V latency/intensity function slope predicted loudness discomfort level, and that using this function, loudness discomfort level could be predicted within 5 dB in a group of human subjects with cochlear impairment (Thornton *et al.*, 1987). This phenomenon was later explored to make the approach feasible for clinical applications (Thornton *et al.*, 1989). Davidson *et al.* (1990) explored the relationship of ABR peak amplitude and loudness in both normal-hearing and hearing-impaired subjects to assess whether such measures might be applied to hearing aid fittings. Darling and Price (1990) covaried click rate and level to produce click stimuli that were judged to be at loudness levels of 70, 80, and 90 phons in normal-hearing listeners. Click stimuli that were judged to be equal in loudness (produced by varying click rate and level) did not produce similar ABR peak latencies or amplitudes, hence demonstrating no clear relationship between loudness and ABR peak latencies or amplitudes. Serpanos *et al.* (1997) assessed the relationship between loudness and the wave V latency/intensity function in subjects with normal hearing, high-frequency sensorineural hearing loss, and flat sensorineural hearing loss. They reported a significant relationship between loudness and wave V latency/intensity function slope in normal-hearing subjects and those with a flat hearing loss, but not those with a high-frequency hearing loss. Collectively, these studies gave a somewhat mixed review about the relationship between click-evoked ABR amplitude or latency and loudness growth. Taken as a whole, the evidence that ABR peak latency or amplitude measures were suitable surrogates for measuring loudness is not compelling.

Two additional studies provided insight into neurophysiological correlates of loudness. Silva and Epstein (2010) estimated loudness growth from ABRs to 1- and 4-kHz tone bursts. By controlling residual noise and parametric curve fitting and using normal hearing young adults, they found that changes in the ABR associated with level (using the cross spectrum of two independent weighted responses) were comparable to loudness growth. Ménard *et al.* (2008) compared the auditory steady state response (ASSR) amplitude growth to loudness growth in normal hearing subjects, using 500 and 2000 Hz carriers. They reported that normalized ASSR amplitude also correlated highly with loudness. These preliminary results using tone burst ABRs and ASSRs evoked by sinusoidally amplitude-modulated sounds suggest that, under certain conditions, AEP measures can be related to loudness. However, considerably more work in this area is

required to determine the extent to which AEP measures are useful surrogates to perceptual loudness measures, especially at frequencies below the region of best hearing sensitivity.

## H. Use of weighting functions in determining the adverse consequences of sound exposure

There are many potential adverse consequences of sound exposure. Governmental agencies often specify policies, procedures, rules, or laws that prescribe what aspects of sound may be permissible in certain situations so as to reduce possible adverse effects of sound exposure (e.g., a weight impact indicator as described in Fig. 1). Where humans are concerned, A-weighting is specified in most of these policies, procedures, rules, or laws (i.e., sound levels are subjected to A-weighting). From a historical perspective, it is unclear whether alternative weighting functions were considered during the development of standards or governmental policies, but it seems likely that the limited frequency weighting-function options available for use in commercially available sound level meters ultimately led to the widespread adoption of A-weighting.

Sound can directly affect the hearing apparatus, but exposure to sound can also elicit a wide range of behavioral consequences in listeners from the pleasant to the unpleasant. Miller (1978a,b,c) and Broadbent (1981) discuss the following possible adverse effects of sound exposure for humans (this list is likely not exhaustive):

- (1) Hearing loss;
- (2) Masking (and disruption in communication);
- (3) Interference with sleep;
- (4) Annoyance;
- (5) Impaired performance;
- (6) Impaired time judgments;
- (7) Effects on other senses;
- (8) Mental disorders;
- (9) Anxiety, distress, and stress;
- (10) Detrimental non-hearing physiological effects;
- (11) Decreased vigilance;
- (12) Impaired memory/recall.

The detrimental noise effects listed above are not entirely independent of each other, and the evidence to support some of the effects of noise range from a high level of evidence (e.g., hearing loss, masking) to more limited evidence (detrimental physiological effects). Obviously, long-term overall health effects from noise exposure are heavily confounded by a myriad of other health-related factors in a given individual's life (comorbid disease, obesity, age, alcoholic beverage consumption, drug use, smoking, risky sexual behavior, etc.). Parsing out the role of noise exposure from the innumerable other factors that can lead to negative health consequences remains a difficult (and perhaps impossible) scientific challenge.

Nevertheless, weighting functions are often used in standards, procedures, policies, and laws that specify measures of sound as they might impact one or more of these adverse consequences. In most cases, the intent is that reliable and valid approaches to sound measurement would

promote some form of alleviation to these adverse consequences for human listeners. It is clear that standards and policies often only assure that everyone makes such measures similarly, albeit imperfectly.

Below, the role of weighting functions for mitigating hearing loss, annoyance, and masking are described because these are the effects for which weighting functions primarily have been developed and used in policy and regulation. That is not to say, that the effects of sound on sleep [e.g., [Pearsons et al. \(1995\)](#)] and several other effects<sup>7</sup> have not also been studied.

### 1. Noise induced hearing loss (NIHL)

There are two main types of NIHL: those caused by acoustic trauma from a very high-level sound of (typically) short duration, and those that occur from exposure to lower level sounds that are presented over substantially longer time periods ([Schuknecht, 1974](#); [Royster, 1996](#)). Later in this article, we present damage risk criteria specified by, e.g., The National Institute of Occupational Safety and Health (NIOSH), where human subjects can be exposed to sounds, e.g., 8 h a day at 90 dBA for 5 days a week, for ten years, that will lead to what “someone” has deemed acceptable hearing loss (in terms of threshold shift in an individual for a given proportion of the exposed human population). These guidelines are, of course, for long duration noise exposures. For acoustic trauma, a brief, high-level signal (for unweighted peak SPL values of, e.g., 125 dB re 20  $\mu$ Pa or higher) can lead to immediate structural damage to the inner ear that causes a sensorineural hearing loss in humans.

Most standards or policies use A-weighted sound level as the basis for evaluating the potential of a sound to produce NIHL from long-term exposures to sound. The use of A-weighting in humans reflects the fact that humans hear best in their mid-frequency region of audibility, with poor thresholds in the lower- and higher-frequency regions of audibility. The determination of sound levels that might produce a hearing loss are derived from experiments of temporary threshold shift (TTS), which is a hearing loss that lasts for minutes to days after a sound exposure, but which eventually recovers to the pre-exposure hearing level. Alternatively, derivations come from population studies of large numbers of subjects exposed on a daily basis, for extended time periods, to fairly stationary noise levels. For ethical reasons, it is not possible to expose human subjects to sounds that might or would produce a permanent hearing loss [i.e., a permanent threshold shift (PTS)], so TTS experiments are used to estimate conditions that might cause PTS [see [Henderson and Hamernik \(1995\)](#)].

There are several variables that could be added to the A-weighting process that might increase the ability of a measure to adequately predict hearing loss following a sound exposure, but the main variables that are currently used are the overall A-weighted SPL and the exposure duration. A wide variety of TTS measures clearly show that above a particular SPL, the combination of exposure SPL and duration determine the amount of hearing loss that will occur (i.e., longer sounds can be lower in SPL than shorter-duration

sounds and produce the same amount of hearing loss). There have been two suggested tradeoffs between level and duration that have been used: the “equal-energy rule” (a 3-dB change in SPL for each doubling of duration) or the “5-dB rule” [a 5-dB change in sound pressure level for each doubling of duration, e.g., as in the United States Code of Federal Regulations (CFR) 1910.95]. Many factors must be considered by damage risk criteria, which include not only risk of hearing loss, but also political, social and economic factors [see [Suter \(1996\)](#) for a discussion of some of these factors]. Most policies specifying the levels of sound that should not be exceeded in order to avoid a hearing loss start with an 8-h exposure (a working day) at 85 or 90 dBA, and then (see [Table II](#)) increase the level as the exposure time decreases, reflecting the tradeoff between level and exposure time. Other corrections are often made in damage risk criteria, such as for type of day or season, spectral content of the sound, type of person being exposed, etc. Thus, agencies like the Occupational Safety and Health Administration (OSHA) of the United States and the U.S. military use A-weighted SPLs based on exposure duration, and other variables, to determine permissible sound exposures for protecting people from hearing loss.

### 2. Annoyance due to sound

Many human activities produce sound that is annoying in some way to some people who are exposed to the sound. Annoyance caused by sound can occur in widely different contexts. However, a physical measure of sound that may be annoying to one set of people (e.g., airplane noise for people not employed by the airplane industry) may not be annoying to another set of people [e.g., people employed by the airplane industry, see [Newman and Beattie \(1985\)](#)].

Transportation noise (e.g., noise around airports and highways) can be annoying to those living, working, and recreating near such transportation locations. Some cities have adopted noise ordinances that restrict the level of sound that can be produced by human activity (e.g., a nightclub) so as to reduce the annoyance to its citizens. Again, A-weighted measures are the basis for evaluating the potential annoyance of sound. One common measure for this purpose is the “equivalent continuous sound level,” or “time-averaged sound level,” symbolized as  $L_{eq}$ . Because the SPL of most noise sources vary over time (e.g., as planes land and take off),  $L_{eq}$  provides an estimated constant sound pressure that would produce the same energy as the fluctuating sound pressure over a given time interval. There are a large number of time-averaged sound pressure measures based on this concept. A few examples are provided here.

*Day-night equivalent level ( $L_{dn}$ ):* A-weighted,  $L_{eq}$  noise level, measured over a 24-h period, with 10-dB added to the levels between midnight and 0700 h and from 2200 h to midnight [as derived from ANSI S1.1-2013 and further described in [Yeager and Marsh \(1998\)](#)].

*Day-evening-night equivalent level ( $L_{den}$ ):* A-weighted,  $L_{eq}$ . Noise level, measured over a 24 h period, with a 10-dB penalty added to the levels between 2200 and 0700 h and a 5 dB penalty added to the levels between 1900 and 2200 h to

reflect people's sensitivity to noise during the night and the evening. This is also referred to as the "community noise equivalent level" [as derived from ANSI S1.1-2013 and further described in Yeager and Marsh (1998)].

*Perceived noise level* ( $L_{PN}$ ): The  $L_{PN}$  is defined as the frequency-weighted sound pressure level obtained by a procedure that combines the sound pressure levels in the 24 one-third-octave bands between 50 Hz and 10 kHz (Harris, 1998). The  $L_{PN}$ , based on equal annoyance contours and a measure called the "noy," was developed to rate the "noisiness" of jets [as derived from Kryter (1959); as shown in Eq. (7)],

$$L_{PN} = 40 + 10 \log_2(\text{noy}). \quad (7)$$

A noy is a linear unit of noisiness or annoyance. It is defined as a unit of perceived noisiness, equal to the perceived noisiness of a one-third octave band of noise centered on 1 kHz and having an A-weighted sound pressure level of 40 dB (ANSI-ASA S3.20-2015), and "n" noys is "n" times as noisy (annoying) as a 1 noy sound.

Thus, a wide variety of situational variables (e.g., the time of day of a sound exposure or the type of airplane) have been combined with A-weighting to establish weighted sound level measurements useful for determining the annoying effects of sound on human activity.

### 3. Masking

All animals, including humans, require a sufficient level of sound above the background noise to perceive sounds relevant to their survival, i.e., they need to determine the sources of sounds (e.g., predators and prey) and to communicate. When relevant sounds are masked, the inability to perform these tasks may be detrimental to a listener. Not all noise masks a signal or a target sound in the same way. Thus, weighting functions are used to determine how effective one sound (the noise) may be in masking another sound (the signal). From the earliest days of auditory physiology experiments, it has been clear that the auditory system can be represented as a series of band-pass filters, where each filter (e.g., each eighth nerve fiber) responds to a limited range of frequencies. In a noisy world, these bandpass filters serve as a way of limiting the effects of high-level broadband noise on the auditory system's response. Perceptually, this means that the effects of noise on a narrow-band (e.g., tonal) signal are limited to a narrow band of masking sounds close to the signal frequency. Herein, this narrow band is referred to as the *critical band*. Although not traditionally considered a weighting function, critical bands are a type of weighting function related to the frequency (spectral) content of sound and they are important in dealing with masking in many different contexts involving human auditory performance.

The critical band was first defined and measured by Fletcher in a series of Bell Lab Reports [see Allen (1996) for a detailed summary of many of Fletcher's contributions, including the critical band]. While the motivation for Fletcher to define the critical band was based on loudness vs intensity summation for tones,<sup>8</sup> the experiments that were

done by Fletcher and others following him usually involved masking. As a result, the application of critical bands is most often associated with masking.

Fletcher's first presentation of his famous "band-widening" experiments was in a tutorial paper [see Allen (1996)]. In the band-widening experiment, the bandwidth of a noise stimulus used to mask a tonal signal that was spectrally centered in the noise was incrementally increased and the signal level required for detection was measured (the spectrum level, i.e., level of sound in a 1-Hz band, of the noise masker remained a constant). Two findings resulting from this and related experiments—critical bands and critical ratios—will be discussed.

*a. Critical band.* As the bandwidth of noise increases, the sound level required for tonal threshold detection also increases. However, at a particular noise masker bandwidth, further increases in bandwidth do not yield additional changes (e.g., increases) in signal detection threshold. This strongly implies that tonal masking is determined by a narrow band of frequencies (up to the bandwidth at which there is no further increase in masked threshold) that is "critical" for masking the tone spectrally centered in the noise. The critical bandwidth is the bandwidth of an assumed internal process (a filter or a process based on the biomechanics of the inner ear) that determines the threshold of a tonal signal spectrally centered in noise. As it turns out, the power within the critical band is proportional to the masked threshold. As the bandwidth of the noise increases toward the critical bandwidth, both the power in the critical band and the masked threshold increase. Once the critical bandwidth is reached, maximum power within the critical band is achieved and no additional masking occurs as the masker bandwidth increases beyond the critical bandwidth.

In the years since Fletcher's conception of the critical band, the use of the critical band and the assumption that the power within the critical band is proportional to masked threshold have been validated many times. More recently, a band-reject noise procedure was used to estimate the critical bandwidth and to derive weighting (filter) functions that might fit the resulting data (Moore, 1989). A particularly useful set (bank) of filters that can account for a wide variety of psychophysical results for human subjects is the Gammatone filter bank originally developed by Patterson *et al.* (1987), but based on auditory nerve data collected by de Boer and de Jongh (1978).

*b. Critical ratio.* In addition to the observation that the power in the critical band was proportional to the tonal masked threshold, Fletcher observed that masked threshold (in units of power) was approximately equal to the power in the critical band. Thus, one can calculate the predicted masked threshold ( $P_t$ ) from knowing the power in the critical band ( $P_{cb}$ ),  $P_t = P_{cb}$ . The predicted masked tonal threshold using the critical band is usually referred to as a "critical ratio." From the critical ratio, one can estimate the critical bandwidth (CBW) by knowing (1) the masked tonal threshold in units of power when the noise was broadband and (2) the noise power per unit bandwidth ( $N_o$ ), i.e.,  $CBW = P_t/N_o$  [see Yost and Shofner (2009) for a full derivation].

The critical ratio approach for estimating the width of the critical band assumes that the masked threshold directly relates to the width of the critical band. The bandwidth of the critical band using the critical ratio is really a proportionality measure, with a proportionality constant,  $K$  [which may be considered a measure of efficiency; see [Patterson et al. \(1982\)](#)]. Thus,

$$P_t = K(P_{cb}) \quad (8)$$

and  $K = 1$  only when the power of the critical band equals the threshold power of the masked signal, an observation that Fletcher made based on his data, but which may not be true for all situations. Indeed, [Yost and Shofner \(2009\)](#) challenged this assumption for animal research and [Patterson et al. \(1982\)](#) have also done so after measuring critical ratios as an estimate of critical bands as a function of aging in humans. Results of these studies have demonstrated that unless one knows the value of the proportionality constant ( $K$ , or the efficiency), one cannot obtain a valid estimate of the critical bandwidth from knowing only the power of the tone at masked threshold. As a result, while the critical ratio might be a helpful predictor of auditory masking, it is no longer recommended as a measure of the critical bandwidth.

In summary, the critical band is a form of frequency (or spectral) weighting that has been very useful in accounting for masking in a wide variety of contexts. There is quite a bit of psychophysical data on the critical band and critical ratio in non-human mammals, but a detailed discussion of these data is not appropriate for this review. The interested reader can find many of these data summarized in [Fay \(1988\)](#), although a great deal of data specific to marine mammals has emerged since this publication ([Erbe et al., 2016](#)).

*c. Masking and the articulation index (AI).* Scientists at Bell Labs developed the “articulation index” (AI) to predict the masking effect of noise on speech that might be transmitted via a telephone. The AI is unique to human speech. The process is contained in the [ASTM \(2008\)](#) standard E1130-08, *Objective Measurement of Speech Privacy in Open Plan Spaces Using Articulation Index*.

While AI applies only to human speech, it is briefly mentioned here because the idea of developing a procedure for predicting the masking effect of noise on other aspects of auditory perception, especially in animal models, may be desirable given the success of the application of AI in telephone communication. Possible dependent variables include sound audibility, sound source location, ability to discriminate sounds varying in one or more physical aspects, sound source identification, or loudness. In humans, it is known that maskers that are presented at low levels (20 or more dB less than that required to mask a signal) still reduce the intelligibility of the signal. Thus, being able to predict the aspects of a sound that can produce masking and by how much would be useful, e.g., for marine mammals. Masked threshold may be less important than the effects of a given masking noise on the perceptual saliency of a given signal needed for the survival of an animal (such as biosonar signals). Such masking studies might be performed using neurophysiological, rather than only perceptual, approaches.

## I. Criticisms of the use of A-weighting

Frequency weighted SPL measures, especially A-weighting, are the most widely used weighting functions. A-weighted sound pressure levels are available on almost every available sound level meter, and it is by far the most common means of measuring sound level. However, A-weighting has received criticisms over the years:

- (1) A-weighting is based on the human 40-phon equal loudness contour, which represents a relatively “soft” sound (i.e., toward the low end of a loudness scale). While soft sounds may be appropriate for dealing with possible interference in telephone communications (the probable original intent of the A-weighting function), they may not be appropriate for annoyance and, especially, for NIHL which occurs at high overall sound levels (see below).
- (2) The underlying loudness data used to determine equal loudness contours are based on statistical averaging of highly variable data. While attempts to provide less variable data have been used in current standards, the equal-loudness contours still represent a statistical average across many listeners who differ in their perception of equal loudness. Thus, A-weighting may not be applicable to any particular listener. Similarly, the A-weighting function fails to capture the uncertainty associated with the variability of the data, as well as the curve fitting approach used to create the function [i.e., there is only one fixed frequency (1000 Hz), and many frequencies for which no loudness data have been collected, necessitating a smooth polynomial fit]. Newer data have changed equal loudness contours from those used to originally create the A-weighting function, and these could potentially address some of the uncertainty. However, the newer loudness functions have not been used to change the A-weighting function; a considerable amount of data historically reported using dBA, and incorporation of the original A-weighting scale in many industry devices, has made such a transition extremely difficult.
- (3) A-weighting is independent of the type of sound to be measured, i.e., the A-weighted SPL can be the same for very different sounds. For instance, it might not be the case that the same weighting is appropriate for determining the annoyance of a sound as compared to a sound’s potential for causing NIHL. Or, perhaps a different frequency weighting is appropriate for determining the annoyance of airplane noise as opposed to noise from a nightclub.
- (4) Multiple factors are predictors of the magnitude of NIHL. These include: exposure to ototoxic drugs, impulse versus continuous noise, short-term versus long-term exposure, whole-body vibration, presence of previous hearing loss from prior noise exposure, eye color, age, race, gender, smoker versus non-smoker, environmental agents, other variables thought to vary from subject to subject (e.g., the acoustic reflex, the efferent auditory system, and noise exposure history), and the occurrence and variability in asymptotic threshold shift,

which is a plateau in threshold under conditions of increasing sound exposure (Melnick and Maves, 1974; Mills, 1976; Ward, 1976; Blakeslee *et al.*, 1978; Humes, 1984; Henderson *et al.*, 1993). Given the multitude of factors potentially affecting NIHL, the likelihood of developing a weighting function universally applicable to human populations seems low. The variability of data in studies of noise-induced TTS [e.g., Mills (1976)], both across subjects, and within-subjects on repeated sessions, suggests that no single weighting function will be able to optimally predict the temporary (or permanent) threshold shift to a given exposure in a given individual on a given day.

- (5) A-weighting is largely derived from studies of human listeners utilizing tonal signals and likely does not fully capture the relationship between complex signals and perceived loudness. It does not account for the frequency spectra of signals and likely underestimates contributions of complex signals across the frequency range of hearing.
- (6) Similar noise exposures (using continuous versus impulse noise) can produce different magnitudes of TTS [i.e., more maximum TTS for impulse noise than continuous noise (Buck, 1982)]. This process is made more complicated by the demonstration that repeated stimulation by the same fatiguing stimulus leads to reduced TTS. This phenomenon is known as sound conditioning or “toughening” (Canlon, 1996; Subramaniam *et al.*, 1996), and may explain some of the within-subject variability in TTS magnitude reported in the literature. Toughening experiments indicate that equal-energy but interrupted (versus continuous) noise exposures may yield differing amounts of TTS or PTS (Hamernik *et al.*, 1974; Lei *et al.*, 1996).
- (7) It has recently been demonstrated in mice that substantial (40+ dB) TTS measured 24 h after noise exposure, but with little or no PTS, caused a delayed but significant degeneration of auditory nerve synapses (Kujawa and Liberman, 2009). This work showed a decrease in ABR peak amplitudes at supra-threshold stimulus levels following the nerve degeneration. Thus, hearing loss may be due to more than hair cell damage, the most often cited cause of hearing loss. Frequency weighting functions, including A-weighting, are normally used as a measure relevant for hearing loss estimation based on the assumption that hair cell loss is the major contributor to hearing loss. Perhaps supra-threshold measures other than those based on loudness might be better in reflecting the consequence for hearing loss based on neural—rather than hair cell—loss.
- (8) It is not clear if A-weighting is the best measure to indicate how one sound masks another sound. It is probably the case that measures that include the critical band would provide a better indicator of “subtotal” masking than an A-weighted SPL.
- (9) Parmanen (2012) provides a comprehensive review of ISO 223–2003, *Acoustics—Normal Equal Loudness Contours* and argues that both mathematical and acoustic

discrepancies in this standard lead to errors in loudness representation.

Despite the criticisms, A-weighted SPL is the most frequently used measure of sound level appearing in guidelines, policies, regulations, rules, and laws dealing with airborne sound. No other frequency weighting function has been used to nearly the same extent as A-weighting. There has not been any widely accepted criticism of the A-weighting function in its broad application, nor has there been any significant evidence that A-weighted SPL measures have led to erroneous (group) estimates of sound levels dealing with hearing loss, annoyance, or other predicted outcomes. However, studies comparing A-weighting vs other weighting functions for determining annoyance, hearing loss, or other auditory perception/performance factors have rarely been carried out. The wide variety of contexts in which A-weighted SPL measures have been used and the lack of evidence for any major problems in its use suggests that either A-weighting is appropriate, or perhaps, that details of the weighting are not crucial as long as spectral components of sound are weighted in some way based on human auditory sensitivity.

#### J. Standards and policies related to weightings for human auditory perception and performance

Standards (such as ANSI, ISO, and IEC) are created by interested groups following strict guidelines, and in the truest sense of the term, are “voluntary” (meaning they have no force of law behind them). A brief list of U.S. standards (ANSI) and international standards (ISO, IEC) related to topics considered in this report can be found in Table I. In contrast to the voluntary nature of standards, legislation can mandate regulations, such as those issued by OSHA, and can require that specific standards be utilized in meeting regulations.

Damage risk criteria for noise exposure (Suter, 1996) addresses the thorny question of exactly how much NIHL is acceptable in a human. This might consider both the magnitude of hearing loss in a given individual and/or the percentage of the population exposed to a given noise (e.g., 90 dBA, 8 h a day, five days a week, for ten years). As already discussed, there is considerable inter-subject variability in terms of PTS arising from a given noise exposure, and hence some individuals (those most susceptible to noise exposure) will develop hearing loss, while others will not. Adding to the complexity of determining what is an acceptable amount of hearing loss, it is far from clear what sort of a threshold elevation should be considered a meaningful hearing loss (>20 dB HL?, >25 dB HL?), and the frequencies chosen to calculate the “average hearing loss” (e.g., 500, 1000, 2000 Hz?) will have a substantial impact in how many individuals exposed to a given noise will become “hearing impaired.”

It is worth comparing two different guidelines/recommendations for workplace noise exposure in the U.S.: the permissible exposure limit of Occupational Safety and Health Administration (2011) and the recommended exposure limit by National Institute of Occupational Safety and

TABLE I. A partial list of ANSI, IEC, and ISO standards relevant to human frequency weighting functions.

---



---

ANSI S1.1 2013, *American National Standard Terminology*.

ANSI S1.4 2014/Part 1/ IEC61672 1:2013, *American National Standard Electroacoustics Sound Level Meters Part 1: Specifications* (Nationally Adopted International Standard).

ANSI S3.4 2007 (R 2012), *American National Standard Procedure for the Computation of Loudness of Steady Sounds*.

ANSI S3.6 2010, *American National Standard Specification for Audiometers*.

ANSI S3.20 1995 (R 2008), *American National Standard Bioacoustical Terminology*.

ANSI S3.44 1996 (R 2006), *American National Standard Determination of Noise Exposure and Estimation of Noise-Induced Hearing Impairment*.

ANSI S12.9 2005/Part 4, *American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 4: Noise Assessment and Prediction of Long-Term Community Response*.

ANSI 12.9 2008/Part 6, *American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 6: Methods for Estimations of Awakenings Associated with Outdoor Noise Events Heard in Homes*.

ANSI 12.19 1996 (R 2011), *American National Standard Measurement of Occupational Noise Exposure*.

ISO 226:2003, *Acoustics—Normal Equal-Loudness-Level Contours*.

ISO 389-7:2005, *Acoustics—Reference Zero for the Calibration of Audiometric Equipment—Part 7: Reference Threshold of Hearing Under Free-Field and Diffuse-Field Listening Conditions*.

ISO 389-8:2004, *Acoustics—Reference Zero for the Calibration of Audiometric Equipment—Part 8: Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Circumaural Earphones*.

ISO 532:1975, *Acoustics—Method for Calculating Loudness Level*.

ISO 9612:2009, *Acoustics—Determination of Occupational Noise Exposure—Engineering Method*.

ISO 13474:2009, *Acoustics—Framework for Calculating Distribution of Sound Exposure Levels for Impulsive Sound Events for the Purposes of Environmental Noise Assessment*.

ISO/TS 15666:2003, *Acoustics—Assessment of Noise Annoyance by Means of Social and Socioacoustic Surveys*.

IEC 616721 Ed. 2.0b: 2013, *Electroacoustics Sound Level Meters Part 1: Specifications*.

---



---

**Health (1998)**. For an 8-h day, OSHA allows an exposure level of 90 dBA, while NIOSH recommends an exposure limit of 85 dBA. OSHA implements a 5 dB increase in exposure limits for each halving of the exposure time, while NIOSH uses a 3 dB exchange rate. These values are compared in Table II. Obviously, the NIOSH recommendations are more conservative and will lead to less hearing loss (i.e., less threshold shift and a smaller proportion of exposed subjects with a hearing loss).

**Suter (1996)** provided a table comparing permissible exposure levels (for an 8 h day), exchange rate, and maximum sound levels allowed in 19 different countries. Although dated, the publication provides interesting facts regarding what has historically been acceptable for noise exposures. For the most part, A-weighted SPL for permissible noise exposures over an 8-h day range from 80 to 90 dBA. Note that all countries included reported acceptable levels in dBA, and not another weighting function. The simple fact that the A-weighting function is in common use across the world argues for its continued use in regulating auditory damage risk, despite concerns about the validity of this particular weighting function for intense sounds. Furthermore, it is worth noting that several countries (China,

Germany, Norway) have a range of acceptable 8-h, A-weighted SPL limits that depend on the activity of the worker. For example, Germany limits the noise SPL to 55 dBA for mentally stressful tasks, and 70 dBA for mechanized office work; for other activities, the upper limit is 85 dBA. The exchange rate is either 3 or 5 dB, depending on the countries regulations (e.g., 5 dB: Brazil, Israel, United States; 3 dB: Australia, Canada, China, United Kingdom). The maximum permissible SPL ranges from 115 to 125 dB, while the maximum pSPL ranges from 130 to 140 dB. Thus, there are surprisingly consistent guidelines across various countries, although how this came to be is historically somewhat murky.

### III. WEIGHTING SOUND: MARINE MAMMAL OVERVIEW

#### A. Marine mammals and noise

Unlike the developmental path of auditory weighting functions in human use, which was initiated to improve telecommunications, the development of weighting functions in marine mammals has largely been driven by environmental concerns related to ocean noise. Questions about the potential effects of intense underwater noise on marine mammals were first substantively raised in the early to mid-1990s, primarily as a result of the Heard Island Feasibility Test (**Munk and Forbes, 1989; Munk, 1991**), subsequent Acoustic Thermometry of Ocean Climate experiments (**Worcester et al., 1999**), and the U.S. Navy John Paul Jones (DDG 53) ship shock trial (**National Marine Fisheries Service, 1994**). These concerns were later strengthened by the development of U.S. Navy low-frequency active sonar systems (**Department of the Navy, 2001b**) and the occurrence of marine mammal stranding events coincident with naval exercises featuring hull-mounted tactical sonars (**Frantzis, 1998**;

TABLE II. A comparison of OSHA and NIOSH allowed/recommended noise exposures for human workers in the United States.

---



---

Time	Level-dBA (OSHA)	Level-dBA (NIOSH)
8 hour	90	85
4 hour	95	88
2 hour	100	91
1 hour	105	94
0.5 hour	110	97
0.25 hour	115	100

---



---

Department of Commerce and Department of the Navy, 2001).

Early efforts to predict and mitigate noise impacts on marine mammals were hampered by the lack of relevant, quantitative information concerning the effects of low-frequency sounds on marine mammals (National Research Council, 1994). Furthermore, several fundamental considerations prevented direct application of human frameworks for noise damage-risk criteria to marine mammals. These included differences in sound propagation in air and water, differences in sound transmission pathways and hearing capabilities in terrestrial and marine mammals, differences in the exposure scenarios of greatest concern for target species, and differences in the ultimate goals of regulatory criteria intended to protect individuals from harm (hearing conservation in humans vs prediction and mitigation of population-level effects on marine mammals). As a result, initial efforts to protect marine mammals from anthropogenic sound relied on simple rules and single-number exposure limits defined by received SPL without regard to the duration, frequency, temporal pattern of the sound exposure, or potential differences in the hearing sensitivities of different marine mammals.

The inadequacy of single number guidelines was soon noted, particularly with respect to the wide variation in hearing sensitivity exhibited by marine mammals across frequencies. Consequently, the need for marine mammal-specific auditory weighting functions to better predict the effects of underwater noise was identified (e.g., Nedwell and Turnpenny, 1998).

## B. Historical development of marine mammal weighting functions

The history of weighting function development and application in marine mammals is far less extensive than that related to humans, and the decisions associated with the evolution of marine mammal weighting functions are much clearer. For these reasons, the historical review of marine mammal weighting functions is more detailed in its relation of data to the development process.

The most appropriate data from which to derive weighting functions for marine mammals would ideally be experimental data relating the effect of interest (e.g., TTS onset) to noise frequency and level for each species of concern. Such direct data have not been historically available. Consequently, over the last ~15 years, a number of different approaches have been taken to derive weighting functions for different species and exposure scenarios. As more data have become available, these efforts have become increasingly more rigorous.

The historical development and application of auditory weighting functions has been heavily influenced by the protected status of marine mammals and the specific language in the governing regulations concerning their welfare. In particular, in the United States, the Marine Mammal Protection Act (MMPA)<sup>9</sup> has had a significant influence on the development of quantitative methods to predict the number of individual marine mammals likely to be injured or disturbed by human-generated noise. In this context, weighting functions have sometimes been applied to predict some auditory effects of noise exposure (such as hearing loss) but have not been applied to non-auditory (physical) effects (e.g., those caused by shock waves), masking, annoyance, or other behavioral effects that are mediated by received sound but which are also heavily dependent on the exposure context. The almost exclusive use of weighting functions for the prediction of NIHL in marine mammals has influenced the methods and data sources from which the functions have been derived, placing heavy emphasis on audiometric data and those relating the occurrence of NIHL to features of the noise exposure.

The following sections review significant steps in the historical development and application of auditory weighting functions to marine mammals. It begins with a review of the audiometric data upon which weighting functions have relied, then presents a chronological review of the specific methods used to derive weighting functions and apply them to marine mammal acoustic impact analyses. Figure 7 provides a timeline of key audiometric studies and applications of weighting function approaches for marine mammals that are discussed in more detail in the following sub-sections.

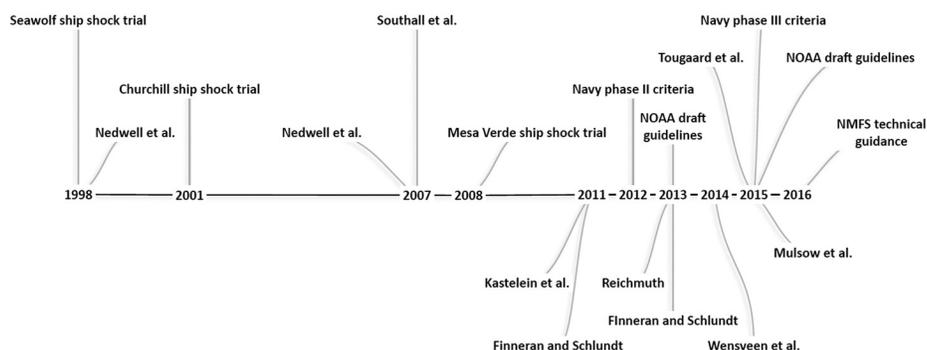


FIG. 7. Timeline of key reports and publications (since 2007) relevant to the development of weighting functions for marine mammals. Items are tied to publication year, and further described in detail in the following. Audiometric studies (Finneran and Schlundt, 2011; Kastelein *et al.*, 2011; Finneran and Schlundt, 2013; Reichmuth, 2013; Wensveen *et al.*, 2014; Mulsow *et al.*, 2015) are discussed in Secs. IV B–IV D. Reports and publications pertaining specifically to weighting function application (Department of the Navy, 1998, 2001a; Southall *et al.*, 2007; National Oceanic and Atmospheric Administration, 2013, 2015; Tougaard *et al.*, 2015) are discussed in Secs. V A–V E.

#### IV. MARINE MAMMAL AUDIOMETRIC DATA

There are 129 species of living marine mammals, which are currently separated into five major groups. These are cetaceans (whales, dolphins, and porpoises), pinnipeds (seals, sea lions, fur seals, and walruses), sirenians (dugongs and manatees), sea otters, and polar bears. Hearing thresholds have been measured in a number of marine mammal species, using psychophysical procedures with trained individuals (e.g., Johnson, 1967) or neurophysiological (AEP) measures with trained or temporarily restrained individuals (e.g., see Dolphin, 2000; Nachtigall *et al.*, 2000). Such hearing data describe the sensitivity of the auditory system as a function of sound frequency, but they have not been collected with standardized approaches as in humans (e.g., MAF standardization), complicating intra- and interspecies comparisons. More advanced audiometric studies related to weighting function development have also been conducted in a few marine mammal species to examine subjective loudness, tonal-detection reaction times (as a proxy for loudness), and NIHL (e.g., Finneran and Schlundt, 2011; Wensveen *et al.*, 2014; Finneran, 2015).

##### A. Auditory thresholds

The first systematic study of marine mammal hearing was performed by Bertel Møhl, who measured psychophysical hearing thresholds and directional hearing in a harbor seal (Møhl, 1964). Since that time, hearing thresholds have been measured underwater in odontocete cetaceans (toothed whales, e.g., dolphins, orcas, porpoises) and sirenians, in air

and underwater in pinnipeds and sea otters, and in air for polar bears (see Erbe *et al.*, 2016). No direct measurements of hearing sensitivity have been made in mysticete cetaceans (baleen whales). Nearly all marine mammals for which data have been obtained have demonstrated ultrasonic hearing capability, with upper-cutoff frequencies extending to ~30–70 kHz for pinnipeds, ~50 kHz for sirenians, and ~150–200 kHz for odontocetes. Underwater hearing thresholds within the frequency range of best hearing sensitivity range from approximately 40 to 60 dB re 1  $\mu$ Pa depending on species. Figure 8 depicts audiograms in six fairly well studied marine mammal species. Note that these audiograms were all obtained using psychophysical methods. The use of AEPs to collect audiograms has increased dramatically over the last decade, but audiograms collected via psychophysical means remain the gold standard. Patterns of frequency-specific sensitivity obtained with AEPs are similar to those obtained with psychophysical measures; however, the absolute sensitivities differ, particularly at the lower frequencies of hearing (Yuen *et al.*, 2005; Finneran and Houser, 2006; Houser and Finneran, 2006; Mulsow and Reichmuth, 2010; Mulsow *et al.*, 2011), thus complicating the integration of AEP and behavioral audiometric data.

Because no direct audiometric data are available for any baleen whale, understanding of hearing in the mysticete group is poor. Inferences of auditory sensitivity are drawn primarily from analyses of sound production, behavioral response studies, and anatomical investigations including efforts to describe auditory structures and to model differential sensitivity as a function of frequency. Findings from all

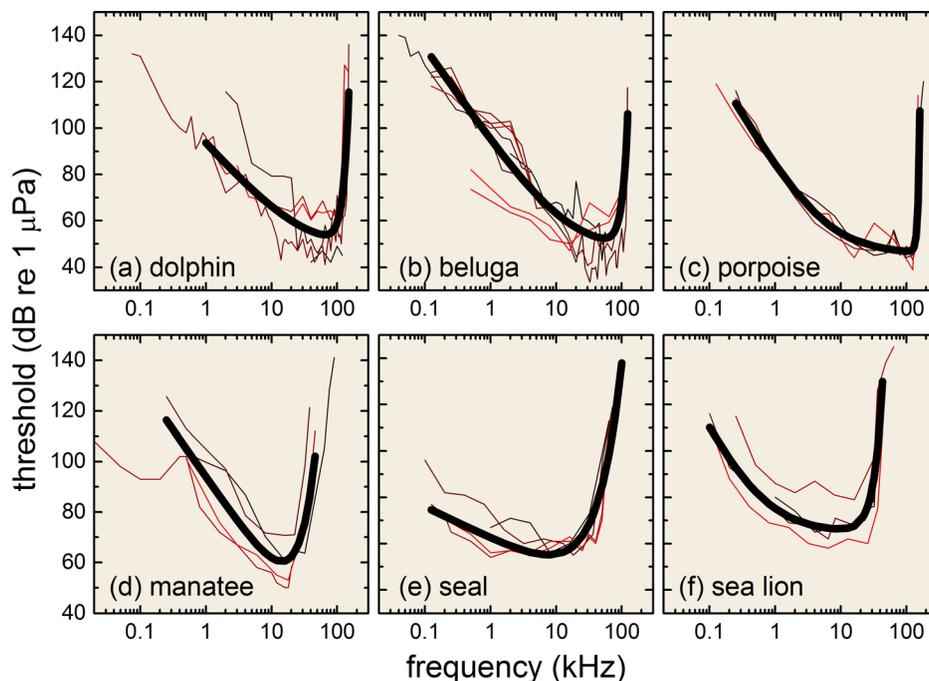


FIG. 8. (Color online) Representative psychophysical audiograms for some of the better studied marine mammals: (a) bottlenose dolphins (Johnson, 1967; Ljungblad *et al.*, 1982; Lemonds, 1999; Schlundt *et al.*, 2008; Finneran *et al.*, 2010a) (b) belugas (*Delphinapterus leucas*) (White *et al.*, 1978; Awbrey *et al.*, 1988; Johnson *et al.*, 1989; Ridgway *et al.*, 2001; Finneran *et al.*, 2005a), (c) harbor porpoises (*Phocoena phocoena*) (Kastelein *et al.*, 2002; Kastelein *et al.*, 2010; Kastelein *et al.*, 2015c), (d) manatees (*Trichechus manatus*) (Gerstein *et al.*, 1999; Mann *et al.*, 2009), (e) harbor seals (*Phoca vitulina*) (Terhune, 1988; Kastelein *et al.*, 2009; Reichmuth *et al.*, 2013), (f) California sea lions (*Zalophus californianus*) (Mulsow *et al.*, 2012; Reichmuth and Southall, 2012; Reichmuth *et al.*, 2013). The thin lines indicate the measured thresholds; thick lines show the best fit of the equation  $T(f) = T_0 + A \log_{10}(1 + F_1/f) + (f/F_2)^B$ , where  $T(f)$  is the threshold at frequency  $f$ , and  $T_0$ ,  $F_1$ ,  $F_2$ ,  $A$ , and  $B$  are fitting parameters to the median threshold at each frequency.

three research areas suggest that mysticetes are low-frequency hearing specialists. At present, given the limited information available, there remains an especially high level of uncertainty in how best to proceed with the development of weighting functions for mysticetes.

## B. Equal loudness contours

Finneran and Schlundt (2011) conducted psychophysical tests with a bottlenose dolphin designed to mimic the approaches used in human loudness comparison tasks (i.e., the comparison method described in Sec. II G 4). These data were then used to estimate equal loudness contours, under the assumption that the auditory systems of humans and dolphins are functionally analogous and therefore dolphins, like humans, may order sound along a dimension that would scale from “soft” to “loud.” The measurements were based on a sound comparison test similar to the loudness comparison test used by Fletcher and Munson (1933) and Robinson and Dadson (1956). In this method, two sequential tones were presented and the dolphin was trained to indicate which tone was perceived to be more intense by emitting one of two vocal responses. One tone (standard tone) was fixed in SPL and frequency, while the other tone (comparison tone) varied in SPL and frequency from one trial to the next. The majority of trials featured tone pairs with identical or similar frequencies but relatively large SPL differences. For these trials it was assumed that the large SPL differences for tones of identical or similar frequencies would be judged in terms of something like a loudness difference for a human, and these trials would enable consistent reward of the dolphin for discriminating a more intense tone from a less intense tone when both tones had the same frequency. A relatively small percentage of trials consisted of “probe” trials, with tone pairs whose presumed loudness relationship was not known by the experimenter—these trials provided the data of primary interest to the study. The dolphin’s responses to the probe trial tone pairings were used to construct psychometric functions describing an assumed loudness relationship between a tone with a particular frequency and sound level and that of a reference tone. These intensity relationships as functions of sound frequency were then used to construct curves that may be analogous to equal loudness contours measured in humans. The data suggested that although these surrogate equal loudness contours in dolphins flatten at higher sound levels (as they do in humans) when either loudness or intensity discrimination is measured as a function of frequency and overall sound level (see above), they generally approximate the audiogram, especially at low overall sound level (Fig. 9). The time required to conduct this study with a single dolphin was extensive—9 months for training and 15 additional months for data collection. As a result, comparable measurements with additional individuals and/or marine mammal species have been considered impractical (Mulsow *et al.*, 2015).

## C. Equal latency contours

As an alternative to direct loudness measurements in non-human animals, several studies have analyzed reaction

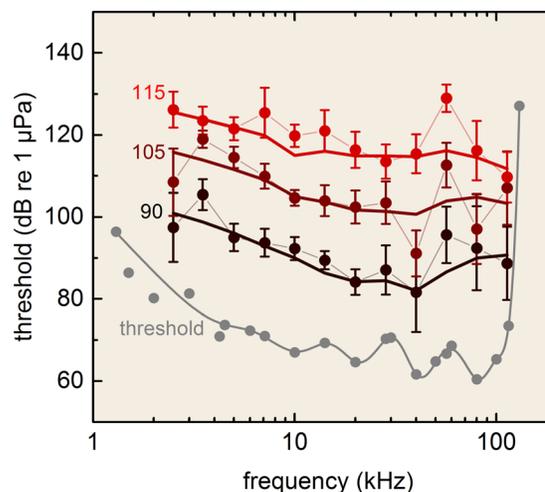


FIG. 9. (Color online) Equal loudness contours derived for a bottlenose dolphin (Finneran and Schlundt, 2011). The loudness level of each contour was defined by the SPL at 10 kHz, in dB re 1  $\mu$ Pa (common SPL provided next to each series). The graph demonstrates, for example, that a  $\sim$ 90 dB SPL tone at 10 kHz had the same perceptual loudness as a 100 dB SPL tone at 2.5 kHz and an 85 dB SPL tone at 40 kHz. Note that at high stimulus amplitudes, perceptual loudness is less influenced by frequency. The dolphin’s audiogram (shown in gray) is provided for reference.

time to estimate loudness. Arguably, the correlation of reaction time and loudness has been demonstrated in humans (Pfungst *et al.*, 1975a; Marshall and Brandt, 1980; Kohfeld *et al.*, 1981; Buus *et al.*, 1982; Wagner *et al.*, 2004), and extension of these measurements to non-human animals has produced equal latency contours that display features seen in human equal loudness contours (Stebbins, 1966; Moody, 1970; Green, 1975; Pfungst *et al.*, 1975a; Pfungst *et al.*, 1975b; Dooling *et al.*, 1978; Ridgway and Carder, 2000; May *et al.*, 2009).

Recent studies have used reaction time measurements to estimate surrogate equal loudness contours in four marine mammal species: the bottlenose dolphin (Mulsow *et al.*, 2015), the harbor porpoise (Wensveen *et al.*, 2014), the harbor seal (Kastelein *et al.*, 2011; Reichmuth, 2013), and the California sea lion (Reichmuth, 2013; Mulsow *et al.*, 2015). The studies featured a similar approach: reaction time was measured in response to tones of various frequencies and SPLs. However, the studies differed in terms of the specific tone sound levels, durations, rise/fall times, and response detection paradigms. In all studies, the reaction time vs SPL data for a given individual were fit with exponential functions and used to estimate equal latency values (see Fig. 10); however, the specific procedures used to create the final contours differed across studies.

For all of the reaction time studies, near-threshold equal latency contours were similar to audiograms, but the contours generally displayed compression at the lower frequencies, similar to trends observed in terrestrial mammals. The dynamic range of the data from Mulsow *et al.* (2015) and Reichmuth (2013) were limited by predictable plateaus in the reaction time vs SPL data; that is, no further reductions in reaction time were observed beyond certain sound levels, limiting the extents to which supra-threshold data could be collected. As a result, the authors of both studies noted that

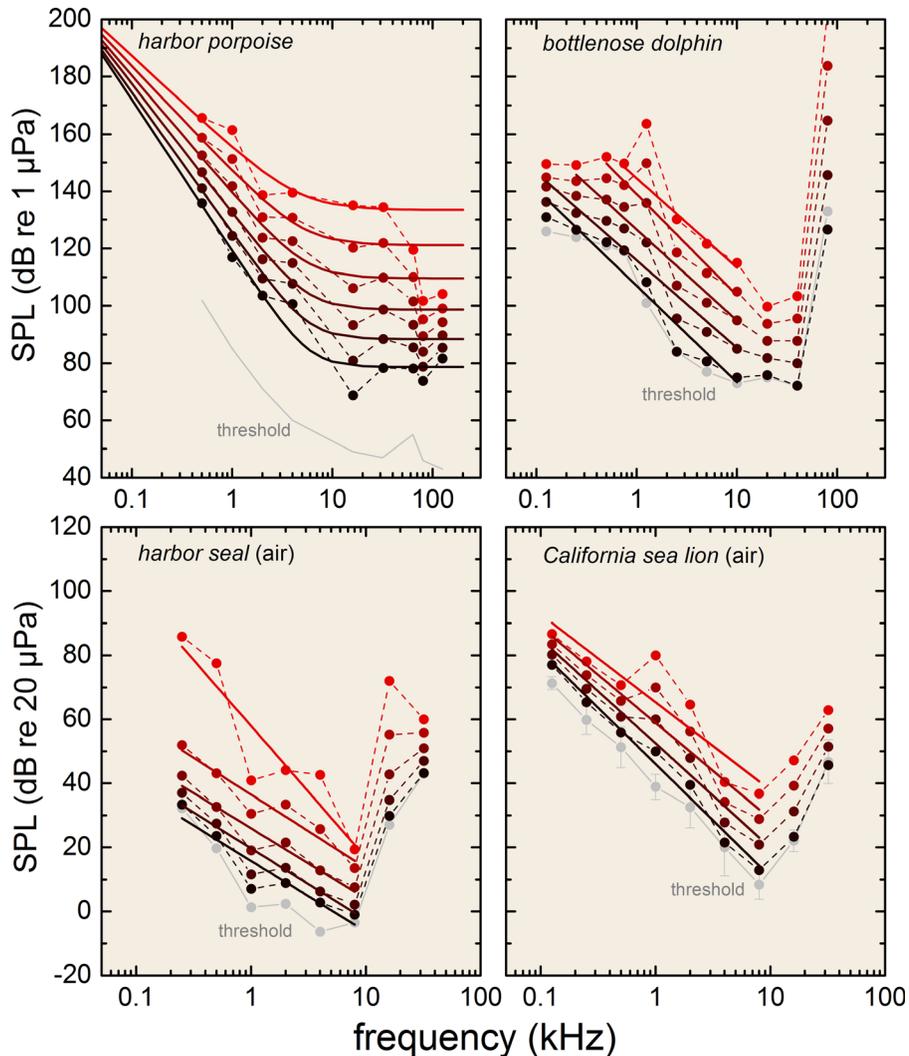


FIG. 10. (Color online) Marine mammal equal latency contours for a harbor porpoise (Wensveen *et al.*, 2014), bottlenose dolphin (Mulsow *et al.*, 2015), harbor seal (Reichmuth, 2013), and California sea lions (Mulsow *et al.*, 2015). The experimental data are based on measures of reaction time to tones of varying frequency and level, with lines connecting similar reaction times across different frequencies. The audiogram for a representative individual of each species (shown in gray) is provided for reference.

equal latency contours appeared to provide limited benefit to predicting the assumed perceptual loudness of higher amplitude sounds (Reichmuth, 2013; Mulsow *et al.*, 2015).

#### D. Noise-induced hearing loss (NIHL)

Exposure to intense noise of sufficient level and duration may result in a NIHL. TTS experiments with marine mammals began in the mid-1990s in response to concerns over the potential effects of naval sonars and underwater explosions on marine mammal hearing (Ridgway *et al.*, 1997). Over the last 20 years, numerous studies have been conducted with dolphins, belugas, porpoises, seals, and sea lions to examine the relationships between TTS and the SPL, duration, frequency, and temporal patterns of noise presentation (reviewed by Finneran, 2015). The primary manner in which TTS data have been used to support the development of weighting functions is to create iso-TTS contours showing how the noise exposure level required to induce a specific amount of TTS varies with frequency. Many studies have used a criterion of 6 dB of TTS (i.e., a 6 dB elevation in post-exposure hearing threshold relative to the pre-exposure hearing threshold) measured 2 to 4 min after exposure to define the “onset” of TTS. This approach is fundamentally different from that used to define human noise damage risk

criteria, which typically seek to limit the risk of PTS. For example, the OSHA criteria are based on limiting the risk of a noise-exposed population developing a “material hearing impairment,” defined as the average threshold across 1, 2, and 3 kHz exceeding 25 dB HL.

Early marine mammal TTS studies used fatiguing noise sources with limited or fixed bandwidth and fixed duration, and induced only small amounts of TTS (e.g., Kastak *et al.*, 1999; Finneran *et al.*, 2000; Finneran *et al.*, 2002; Finneran *et al.*, 2003; Nachtigall *et al.*, 2003; Nachtigall *et al.*, 2004). In the only study from this time period to examine the effects of exposure frequency, Schlundt *et al.* (2000)<sup>10</sup> exposed dolphins and belugas to 1-s tones at frequencies of 0.4, 3, 10, 20, and 75 kHz. The resulting threshold shifts, measured through behavioral methods 1 to 3 min post-exposure, were generally small and there were only small differences in TTS onset as a function of exposure frequency. The minimum exposure SPLs resulting in TTS  $\geq 6$  dB were  $\sim 194$ , 192, 193 dB re  $1 \mu\text{Pa}$  at 3, 10, and 20 kHz, respectively; the mean value was 195 dB re  $1 \mu\text{Pa}$  for a 1-s tone (195 dB re  $1 \mu\text{Pa}^2 \text{ s}$ ). Because the exposure durations and frequency ranges were representative of the most powerful naval sonars, data from Schlundt *et al.* (2000) had a large influence on the development of acoustic impact thresholds and weighting functions over the next several years.

Between 2000 and 2004, three behavioral studies were conducted to examine the effects of discrete underwater impulsive sounds (as opposed to more continuous broadband or tonal sounds) on dolphins, belugas, and California sea lions. [Finneran \*et al.\* \(2000\)](#) exposed dolphins and a beluga to single impulses from an array of underwater sound projectors designed to produce pressure signatures resembling underwater explosions, but found no TTS after exposure to the highest level the device could produce (unweighted SEL = 179 dB re 1  $\mu\text{Pa}^2$  s). Similarly, no TTS was found in two California sea lions exposed to single impulses from an arc-gap transducer with unweighted SELs of 161 to 163 dB re 1  $\mu\text{Pa}^2$  s ([Finneran \*et al.\*, 2003](#)). [Finneran \*et al.\* \(2002\)](#) reported TTSs of 6 and 7 dB in a beluga exposed to single impulses from a seismic water gun (unweighted SEL = 186 dB re 1  $\mu\text{Pa}^2$  s, pSPL = 224 dB re 1  $\mu\text{Pa}$ ).

Over the next decade, TTS testing was primarily focused on the relationships between SPL, duration, and duty cycle for mid-frequency (1–10 kHz) tonal or broadband exposures ([Finneran \*et al.\*, 2005b](#); [Kastak \*et al.\*, 2005b](#); [Kastak \*et al.\*, 2007](#); [Mooney \*et al.\*, 2009a](#); [Mooney \*et al.\*, 2009b](#); [Finneran \*et al.\*, 2010a,b](#)). These behavioral studies demonstrated that TTS from exposures with different SPLs and durations was correlated with SEL; however, for exposures with the same SEL, longer duration exposures would tend to produce more TTS, consistent with findings from human studies. TTS from intermittent exposures was also shown to be lower than that from continuous exposures with the same SEL, but higher than that from a single exposure with the same sound pressure level, i.e., TTS accumulated over multiple, identical exposures but the recovery that occurred between pulses lowered the resulting TTS, so it would be less than that predicted from a single, continuous exposure with the same SEL. During this time there were no systematic efforts to explore the influence of exposure frequency on TTS in marine mammals.

[Lucke \*et al.\* \(2009\)](#) reported the first TTS data for the harbor porpoise, using methods similar to those developed by [Nachtigall \*et al.\* \(2004\)](#) for dolphins. [Lucke \*et al.\* \(2009\)](#)

obtained AEP measurements to determine hearing thresholds in a harbor porpoise before and after it was exposed to air gun impulses at various ranges. The maximum TTS ranged from  $\sim 7$  to 21 dB after exposure to single impulses with unweighted SEL of 165 dB re 1  $\mu\text{Pa}^2$  s and pSPL of 196 dB re 1  $\mu\text{Pa}$ . These data were controversial at the time, since the exposure levels were substantially smaller than those required for TTS in a beluga exposed to watergun impulses (unweighted SEL = 186 dB re 1  $\mu\text{Pa}^2$  s, peak SPL = 224 dB re 1  $\mu\text{Pa}$ , [Finneran \*et al.\*, 2002](#)); however, subsequent studies have confirmed that porpoises appear to be more susceptible to NIHL than dolphins and belugas (e.g., [Kastelein \*et al.\*, 2012b](#)).

Utilizing AEP methods, [Finneran \*et al.\* \(2007\)](#) reported relatively high TTS growth rates and long recovery times in a dolphin exposed to 20-kHz tones, compared to what had been previously measured for the same individual after exposure to 3-kHz tones. In 2010, large differences were reported between TTS growth curves in the same dolphin after 3- and 20-kHz exposures with the same duration ([Finneran and Schlundt, 2010](#)). These results called into question earlier data suggesting little change in TTS onset from 3 to 20 kHz and demonstrated the need for systematic measurement of TTS onset as a function of exposure frequency.

Since 2010, TTS growth curves (see Fig. 11) have been measured at multiple frequencies using psychophysical methods with dolphins and porpoises (e.g., [Kastelein \*et al.\*, 2012b](#); [Finneran and Schlundt, 2013](#); [Kastelein \*et al.\*, 2013b](#); [Kastelein \*et al.\*, 2014b](#); [Kastelein \*et al.\*, 2015b](#)). Figure 12 shows the relationship between exposure frequency and the SEL required for the onset of TTS (6 dB) for bottlenose dolphins, a harbor porpoise, harbor seals, and a California sea lion. In addition, a number of neurophysiological studies have demonstrated significant differences in TTS after band-limited noise exposures with different center frequencies ([Popov \*et al.\*, 2011](#); [Popov \*et al.\*, 2013](#)); however, these studies have not been focused on identifying the “onset” of TTS and have not reported growth curves below  $\sim 10$  dB of TTS.

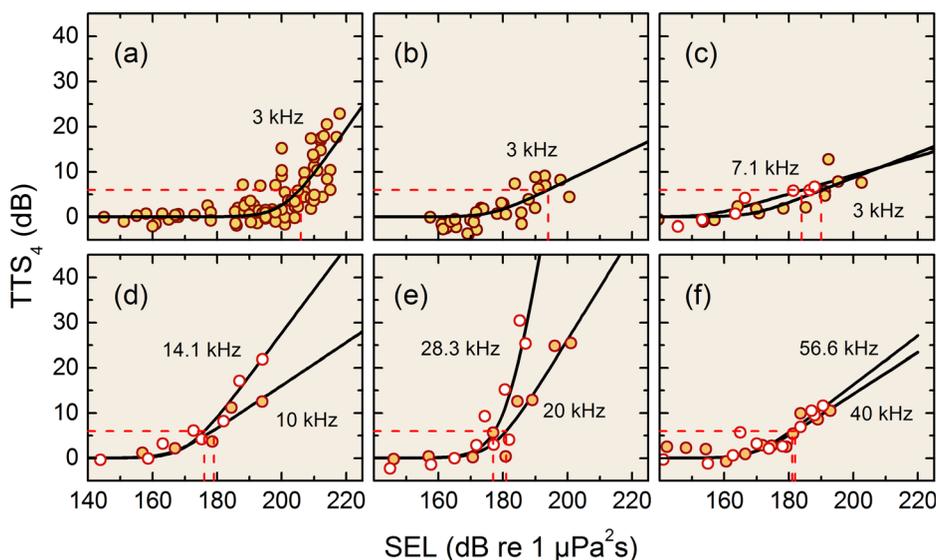


FIG. 11. (Color online) Example TTS growth curves at different sound frequencies for bottlenose dolphins tested with psychophysical methods. Exposures consisted of single, 16-s tones. Hearing was tested 1/2-octave above the exposure frequency. Panels (a), (c), (d), and (e) show data from the same dolphin; panels (b) and (f) show data from a different dolphin. The steepness of the growth curves indicates greater susceptibility to TTS at certain frequencies.

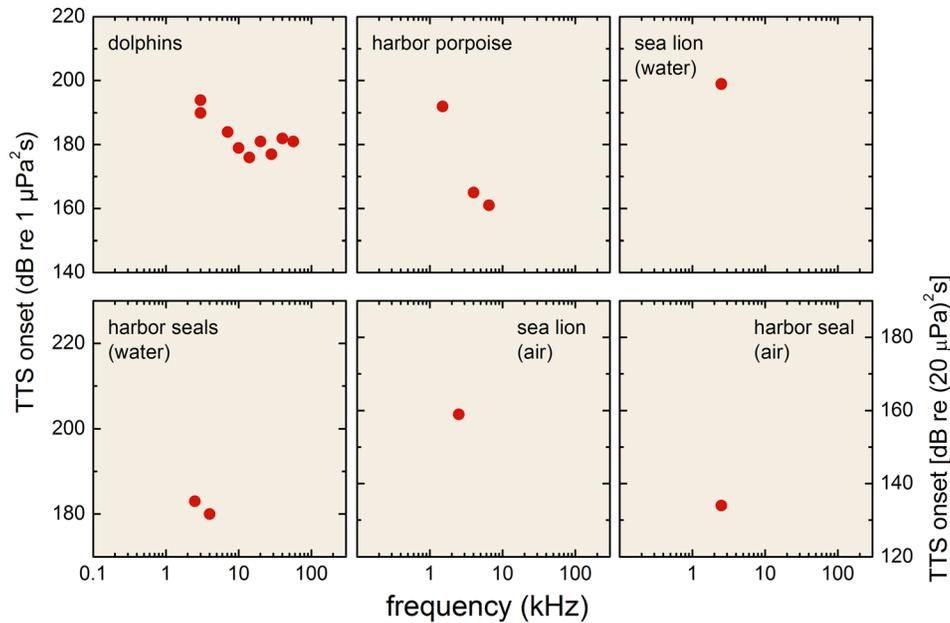


FIG. 12. (Color online) Sound exposure level (SEL) required to induce 6 dB of behaviorally measured TTS at various frequencies in bottlenose dolphins (Finneran *et al.*, 2010a; Finneran and Schlundt, 2013), a harbor porpoise (Kastelein *et al.*, 2012b; Kastelein *et al.*, 2014a; Kastelein *et al.*, 2014b), a California sea lion (Kastak *et al.*, 2005b; Kastak *et al.*, 2007), and harbor seals (Kastak *et al.*, 2004; Kastak *et al.*, 2005a; Kastak *et al.*, 2005b; Kastelein *et al.*, 2012a). Note that the right Y-axis corresponds to the aerial studies. For broadband exposures, frequency is approximated by the center of the band. The lowest onset TTS values are shown for each subject and frequency. Onset values were determined by interpolating within the TTS growth curves (e.g., Fig. 11) showing TTS as a function of SEL, for only those datasets with TTS values bracketing 6 dB, i.e., no extrapolation was performed.

## V. MARINE MAMMAL AUDITORY WEIGHTING FUNCTIONS

Initial efforts to derive meaningful auditory weighting functions for marine mammals were hampered by the limited and sometimes conflicting data describing the effects of noise at various frequencies. As more data have become available, approaches used to define weighting functions have become more sophisticated (i.e., more complex). Key experimental approaches and analytical applications used thus far in the development of weighting functions for marine mammals are summarized in Fig. 13.

### A. Sensation level-based functions (1998 to present)

The most straightforward approach for developing auditory weighting functions is to weight the predicted noise exposure by the animal's relative hearing sensitivity. In essence, this approach inverts the auditory threshold curve to define the weighting function shape, so that the noise at each

frequency is weighted by the amount it exceeds the animal's hearing threshold at the same frequency. This means the weighted SPLs essentially represent the sensation level of the noise. This approach is equivalent to the “dBht” method proposed by Nedwell and Turnpenny (1998) and Nedwell *et al.* (2007); however, it should be noted that sensation level and perceptual loudness are not equivalent, except at threshold. Therefore, weighting noise exposures by auditory sensitivity does not provide a measure of perceived loudness.

Direct experimental data to support the use of sensation level-based weighting functions have been limited. It seems clear that noise below an animal's hearing threshold cannot result in adverse auditory effects or behavioral disturbance, but that does not necessarily mean adverse effects simply scale with sensation level for SPLs well above threshold. TTS experiments with pinnipeds have provided some support for sensation level-based weighting to equate exposure levels across individuals at the same frequency (Kastak *et al.*, 1999), or to compare the effects of airborne and

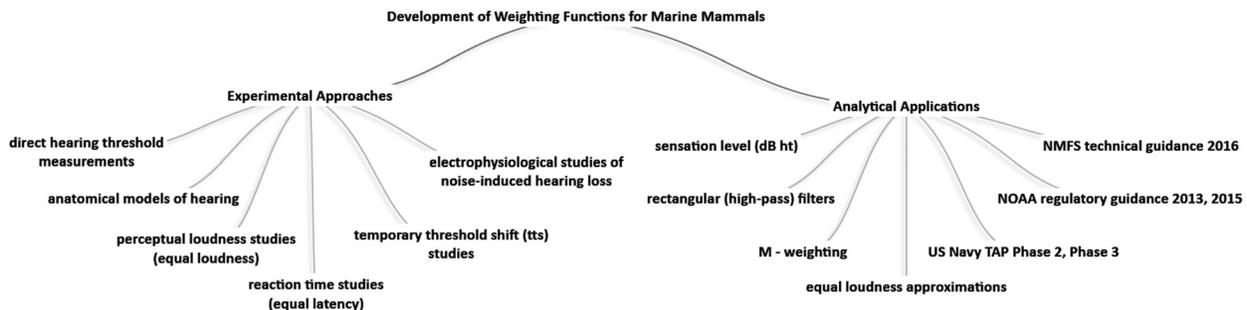


FIG. 13. Experimental approaches (discussed in Sec. IV) providing data to support the development of weighting functions of marine mammals are indicated on the left, while analytical frameworks derived from these data are shown on the right (this section).

underwater noise at the same frequency on the same individual (i.e., noise exposures of equal durations could be equated in terms of SL, irrespective of medium, [Kastak \*et al.\*, 2005b](#)). However, the relationships did not always hold, and [Kastak \*et al.\* \(2007\)](#) suggested that sensation level-based weighting may have a limited frequency range over which it is applicable.

Adoption of sensation level-based weighting was countered by the early experimental data from [Schlundt \*et al.\* \(2000\)](#), which showed only small changes in TTS onset in dolphins in the frequency range from 3 to 20 kHz, a range over which thresholds changed significantly. There was also widespread concern that sensation level-based functions would underestimate the effects of noise at low frequencies, where odontocete thresholds were poor and increased with decreasing frequency at a rate of  $\sim 30$  dB/decade (e.g., [Fig. 8](#)). More recent TTS data ([Kastelein \*et al.\*, 2012a](#); [Finneran and Schlundt, 2013](#); [Kastelein \*et al.\*, 2014a](#); [Kastelein \*et al.\*, 2014b](#)) have provided greater support for sensation level-based approaches at frequencies below the region of best sensitivity. However, neurophysiological TTS measurements with belugas and porpoises ([Popov \*et al.\*, 2011](#); [Popov \*et al.\*, 2013](#)) have shown exposure frequencies of  $\sim 11$  to 32 kHz—below the most sensitive frequency range—to produce larger amounts of TTS compared to exposures at higher frequencies, where hearing thresholds were lower.

Despite these issues, sensation level-based weighting functions have been applied in some situations, such as predicting the effects of low-frequency shipping noise on porpoises ([Terhune, 2013](#)). Some recent weighting function proposals have also advocated a sensation level-based approach (e.g., [Tougaard \*et al.\*, 2015](#)), or used the audiogram as a starting point for weighting function derivation ([Finneran, 2016](#)).

## B. Rectangular (high-pass) filters (1998–2001)

The first well-documented application of auditory weighting functions to predict the effects of noise on marine mammals occurred in conjunction with U.S. Navy testing to determine the ability of vessels to withstand explosive impacts ([Fig. 14](#)). Acoustic impact analyses for the USS SEAWOLF ([Department of the Navy, 1998](#)) and USS WINSTON CHURCHILL ([Department of the Navy, 2001a](#)) featured criteria for marine mammal injury based on the highest SEL in any 1/3-octave frequency band, but only considered bands  $\geq 100$  Hz for odontocetes and those  $\geq 10$  Hz for mysticetes (there were no sirenians or marine carnivores in the proposed operating areas). This approach essentially utilized a weighting function based on a rectangular, high-pass filter with a cutoff frequency of 10 Hz for mysticetes and 100 Hz for odontocetes. Above the cutoff frequency, the function was flat (i.e., no weight was applied, all noise energy was considered equally hazardous). Below the cutoff frequency, all noise energy was ignored. The specific cutoff frequencies were based on the  $\sim 70$  dB sensation level for dolphins occurring at  $\sim 100$  Hz (based on available audiometric data) and anticipated

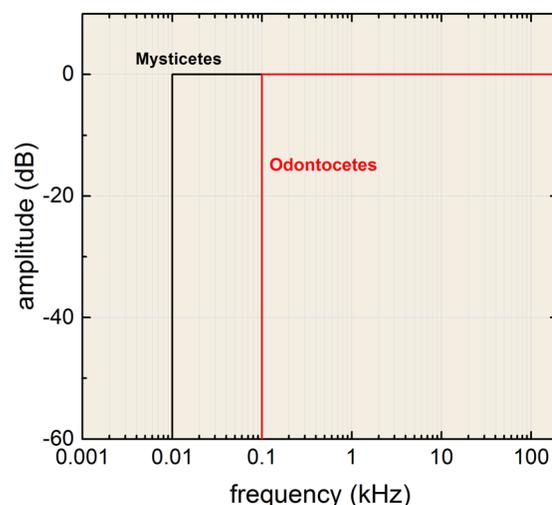


FIG. 14. (Color online) Rectangular filter utilized for U.S. Navy shock trials of the USS SEAWOLF ([Department of the Navy, 1998](#)) and USS WINSTON CHURCHILL ([Department of the Navy, 2001a](#)). The filter used was essentially a weighting function for mysticetes that excluded received noise energy below 10 Hz; the function for odontocetes excluded received noise energy below 100 Hz.

thresholds for mysticetes extending below those of dolphins ([Department of the Navy, 2001a](#)).

These rectangular filters were used in conjunction with a weighted TTS threshold of 182 dB re  $1 \mu\text{Pa}^2 \text{ s}$ . This number was derived using the minimum SEL necessary for TTS  $\geq 6$  dB reported by [Ridgway \*et al.\* \(1997\)](#) for bottlenose dolphins: 192 dB re  $1 \mu\text{Pa}^2 \text{ s}$ . The 192 dB re  $1 \mu\text{Pa}^2 \text{ s}$  value was then lowered by 10 dB based on an estimated  $\sim 100$  ms temporal integration time for dolphins ([Department of the Navy, 1998](#)). An unweighted peak SPL threshold of 83 kPa (12 psi) was also used to predict TTS; if either TTS threshold was exceeded for a cetacean in the modeled area, TTS was assumed ([Department of the Navy, 1998](#)). Similar “dual criteria” methodologies have been widely used for marine mammal auditory impact predictions, especially for impulsive noise exposures.

## C. “M-weighting” (2007)

The first broadly applied marine mammal weighting functions were developed by [Southall \*et al.\* \(2007\)](#) and referred to as the M-weighting functions. At this time, there were only limited TTS data for marine mammals, no equal loudness or latency contours existed for marine mammals, and the existing marine mammal TTS data showed little variation with frequency (at least between 3 and 20 kHz). For these reasons, [Southall \*et al.\* \(2007\)](#) based their proposed weighting functions on the shape of the human “C-weighting” function, with the parameters adjusted so the weighting function shape better matched the known or suspected hearing ranges for various groupings of marine mammals.

The functions were described by the equation

$$W(f) = k + 20 \log_{10} \left[ \frac{b^2 f^2}{(a^2 + f^2)(b^2 + f^2)} \right], \quad (9)$$

TABLE III. Parameters for the “M-weighting” functions and associated acoustic impact thresholds defined by Southall *et al.* (2007) for five categories of marine mammals. TTS and PTS thresholds are in terms of weighted SELs, with units dB re  $1 \mu\text{Pa}^2 \text{s}$  in water and dB re  $(20 \mu\text{Pa})^2 \text{s}$  in air. Note that the mysticetes are identified as LF, most odontocetes as MF, high-frequency odontocetes as HF, and that pinnipeds have separate parameters for waterborne and air-borne noise exposures. Some marine mammals (sirenians, sea otters, walrus, and polar bear) are not included.

Species group	Parameters for Eq. (9)			Non-impulse SEL threshold [dB re $1 \mu\text{Pa}^2 \text{s}$ or dB re $(20 \mu\text{Pa})^2 \text{s}$ ]		Impulse SEL threshold [dB re $1 \mu\text{Pa}^2 \text{s}$ or dB re $(20 \mu\text{Pa})^2 \text{s}$ ]	
	$k$	$a$ (Hz)	$b$ (Hz)	TTS	PTS	TTS	PTS
LF—Low-frequency cetaceans	0	7	22 000	195	215	183	198
MF—Mid-frequency cetaceans	0	150	160 000	195	215	183	198
HF—High-frequency cetaceans	0	200	180 000	195	215	183	198
PW—Pinnipeds in water	0	75	75 000	183	203	171	186
PA—Pinnipeds in air	0	75	30 000	131	144	129	144

where  $f$  is the frequency (Hz),  $W(f)$  is the weighting function amplitude (dB) as a function of frequency,  $a$  and  $b$  are constants related to the upper and lower hearing limits, respectively, and  $k$  is a constant used to normalize the equation at a particular frequency. Unique functions were defined for five groups of cetaceans and pinnipeds: low-frequency cetaceans (mysticetes), mid-frequency cetaceans (e.g., delphinids, beaked whales), high-frequency cetaceans (e.g., porpoises), pinnipeds in air, and pinnipeds in water. Specific values for the constants  $a$  and  $b$  are given in Table III (Southall *et al.*, 2007) and the bandpass shape of the resultant weighting functions are shown in Fig. 14. Consistent with this approach, the U.S. Navy later utilized a variant of M-weighting for acoustic effects analyses for the MESA VERDE ship shock trial, with the lower cutoff frequency,  $a$ , for mysticetes increased from 7 to 12 Hz (Department of the Navy, 2008).

Southall *et al.* (2007) also proposed revised thresholds for auditory injury of marine mammals based on the onset of PTS, which was estimated to occur after any exposure producing a TTS (measured  $\sim 2$ – $4$  min after exposure)  $\geq 40$  dB. Exposure levels sufficient to produce threshold shifts of 40 dB were estimated from TTS thresholds and TTS growth rates from available marine and terrestrial mammal data (see Southall *et al.*, 2007). TTS thresholds utilized dual criteria based on M-weighted SEL (Table III) and unweighted pSPL. Separate thresholds were proposed for impulsive and non-impulsive exposures based on similar logic.

For non-impulsive noise exposures, the weighted SEL TTS thresholds for cetaceans were based on data for dolphins and belugas from Schlundt *et al.* (2000). For pinnipeds, the in-water threshold was based on harbor seal data from Kastak *et al.* (1999, 2005b) and the in-air threshold was based on harbor seal data from Kastak *et al.* (2004). For impulsive exposures, the weighted SEL-based TTS thresholds for cetaceans were based on TTS data from a beluga exposed to water gun impulses (Finneran *et al.*, 2002). For pinnipeds in water, the SEL-based impulsive threshold was estimated using a multi-step extrapolation process (Southall *et al.*, 2007). The weighted TTS threshold for pinnipeds in air was based on unpublished data for harbor seals exposed to impulse noise (Southall *et al.*, 2007).

The M-weighting functions were nearly flat between the lower and upper cutoff frequencies ( $a$  and  $b$ , respectively)

specified in Table III (Fig. 15); at the time this was believed to result in conservative criteria in the absence of data. However, the weighted exposure thresholds utilized with each function were based on the available marine mammal TTS data (Fig. 16), which primarily existed for frequencies below 10 kHz. As a result, the intentionally broad, flat nature of the functions resulted in an under-estimate of the effects of noise at frequencies within the range of best hearing sensitivity, above the frequency range of the existing TTS data (i.e., frequencies above 10 kHz). This limitation in the M-weighting functions became clear once marine mammal TTS data collection occurred at higher frequencies (Finneran and Schlundt, 2010; Finneran and Schlundt, 2013; Kastelein *et al.*, 2014a; Kastelein *et al.*, 2014b; Kastelein *et al.*, 2015b), suggesting differential vulnerability to hearing loss as a function of exposure frequency.

#### D. Dolphin equal loudness-based functions (2011)

Finneran and Schlundt (2011) estimated equal loudness contours using procedures similar to those used to derive human equal loudness contours (e.g., Suzuki and Takeshima, 2004). The resulting contours were similar in shape to the

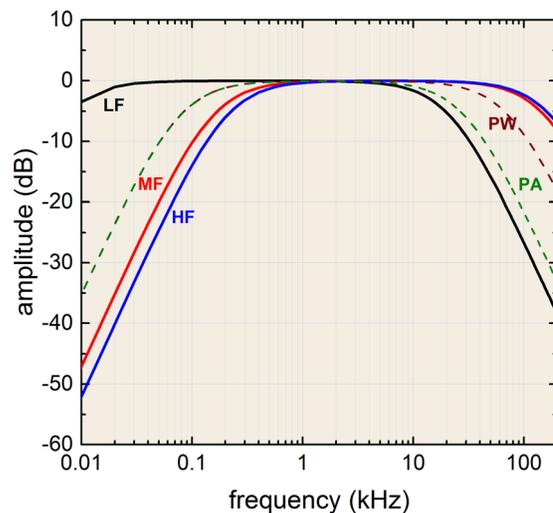


FIG. 15. (Color online) Marine mammal auditory weighting functions proposed by Southall *et al.* (2007). LF—low-frequency cetacean, MF—mid-frequency cetacean, HF—high-frequency cetacean, PW—pinnipeds in water, PA—pinnipeds in air. Amplitude (dB) refers to the amount by which a given predicted noise exposure is weighted for the marine mammal group.

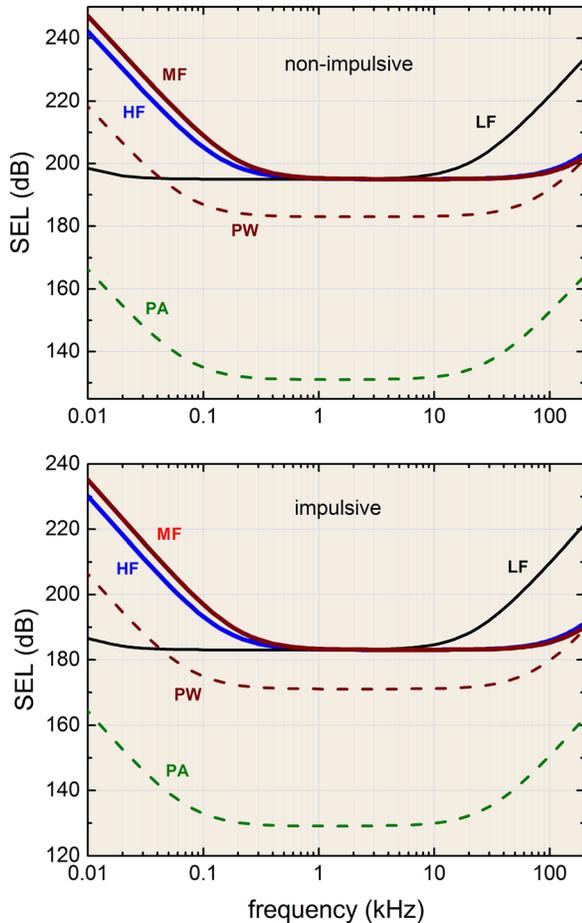


FIG. 16. (Color online) TTS exposure functions based on the M-weighting functions for (upper) non-impulsive and (lower) impulsive TTS impact thresholds (Southall *et al.*, 2007). LF—low-frequency cetacean, MF—mid-frequency cetacean, HF—high-frequency cetacean, PW—pinnipeds in water, PA—pinnipeds in air. Units for SEL are dB re  $1 \mu\text{Pa}^2 \text{s}$  underwater (groups LF, MF, HF, PW) and dB re  $(20 \mu\text{Pa})^2 \text{s}$  in air (group PA).

audiogram and became flatter as tone SPL increased, as expected, providing confidence that the method provided data somewhat similar to that available for human listeners. Equation (9) was fit to the estimated equal loudness contours, providing a set of auditory weighting functions (the “EQL weighting functions”). Three weighting functions were derived, based on the estimated equal loudness contours passing through 90, 105, and 115 dB re  $1 \mu\text{Pa}$  at 10 kHz. As shown in Fig. 17, the shapes of these three weighting functions, especially the 90 dB SPL (re  $1 \mu\text{Pa}$ ) function, agreed closely with independent measurements of TTS in dolphins over the frequency range 3 to 56 kHz (Finneran and Schlundt, 2013); however, the frequency range for which the data existed (2.5–113 kHz) prevented estimates for the weighting function shape at lower frequencies.

### E. U.S. Navy “phase 2” weighting functions (2012)

As part of the U.S. Navy’s Tactical Training Theater Assessment and Planning (TAP) program, acoustic effects analyses are conducted to estimate the potential effects of Navy activities that introduce high levels of sound or

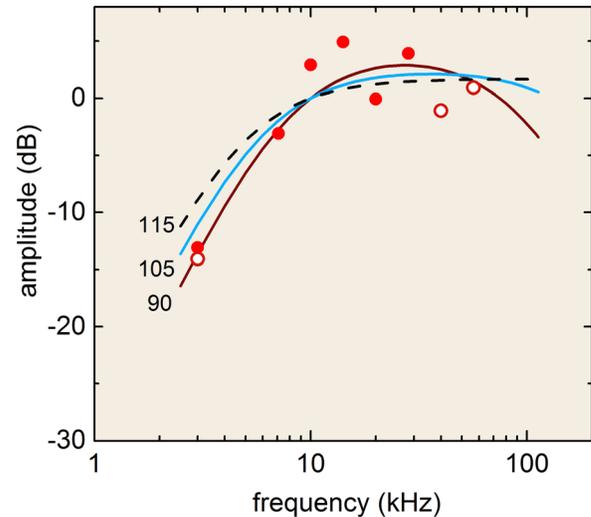


FIG. 17. (Color online) Comparison of dolphin auditory weighting functions (lines) and relative susceptibility to noise (symbols) measured in bottlenose dolphins. The numeric values indicate the equal loudness contour upon which each weighting function was based (Finneran and Schlundt, 2011). The relative susceptibility data were obtained from experimental studies of TTS in dolphins (Finneran and Schlundt, 2013). Source data are reported individually in Figs. 10 and 13.

explosive energy into the marine environment. For phase 2 of the TAP program, two types of weighting functions were defined for select marine mammal groupings, referred to as the “type I” and “type II” weighting functions (Finneran and Jenkins, 2012). For auditory effects (TTS and PTS), the type I functions were used for all non-cetaceans and the type II functions were used for cetaceans. For behavioral effects from non-impulsive sources, type I functions were used except for species considered especially sensitive (e.g., harbor porpoises, beaked whales), for which no weighting was applied. The weighting functions were used in conjunction with weighted SEL thresholds and, for explosives and other impulsive sources, unweighted pSPL thresholds for TTS and PTS.

Type I weighting functions were similar to the M-weighting functions [Eq. (9)], with two parameters ( $a$ ,  $b$ ) to define the lower- and upper-cutoff frequencies and one parameter ( $k$ ) to define the amplitude of the flat portion of the curve (Table IV, Fig. 18). As with the M-weighting functions, the cutoff frequencies were based on the known or estimated hearing range for each species group.

Type II weighting functions modified the M-weighting or type I functions (Fig. 18) by including a region of increased amplitude (increased susceptibility) based on the equal loudness-based weighting functions derived by Finneran and Schlundt (2011) for dolphins (Fig. 19). Type II functions were only derived for the cetaceans, because the underlying data necessary for the functions were only available for bottlenose dolphins (mid-frequency cetaceans) and extrapolation beyond the cetacean group was considered questionable. Although TTS data existed at that time for three pinniped species (harbor seal, California sea lion, northern elephant seal), most exposures consisted of octave-band noise centered at 2.5 kHz; thus, data were insufficient to either derive weighting functions in a manner analogous

TABLE IV. Parameters for the type I mammal weighting functions used for U.S. Navy TAP Phase 2 (Finneran and Jenkins, 2012). Note that the species groupings are similar to those used initially by Southall *et al.* (2007) with the further splitting of the marine carnivores into phocid seals and all other marine carnivores.

Species group	Parameters for Eq. (9)		
	$k$	$a$ (Hz)	$b$ (Hz)
LF—Low-frequency cetaceans	0	7	22 000
MF—Mid-frequency cetaceans	0	150	160 000
HF—High-frequency cetaceans	0	200	180 000
PW—Phocids (in water), Sirenians	0	75	75 000
PA—Phocids (in air)	0	75	30 000
OW—Otariids, odobenids, mustelids, ursids (in water)	0	100	40 000
OA—Otariids, odobenids, mustelids (in air)	0	100	30 000

to that used for mid-frequency cetaceans or to verify the effectiveness of extrapolations from the mid-frequency cetacean group.

Type II functions were defined using two component curves: a relatively broad curve based on the type I weighting function and a sharper curve based on the EQL-based weighting function. At each frequency, the amplitude of the weighting function was defined using the larger amplitude from the two component curves (Fig. 19). In practice, the type I component dominated below some frequency, denoted as the “inflection point” frequency, and the equal loudness-based component dominated above the inflection point.

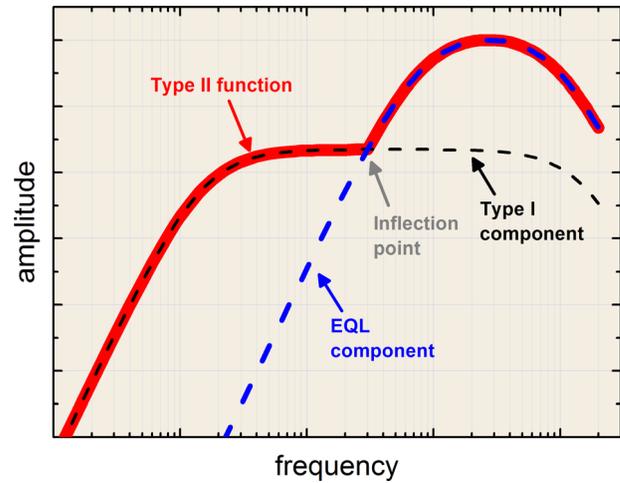


FIG. 19. (Color online) Illustration of the type II weighting function concept. Below the inflection point frequency, the type II weighting function matches the shape of the type I function (see Fig. 18). Above the inflection point, the type II function matches the EQL-based weighting function (see Fig. 17).

The type II weighting functions were mathematically defined as

$$W_{II}(f) = \text{maximum}\{G_1(f), G_2(f)\}, \quad (10)$$

where  $W_{II}(f)$  is the weighting function amplitude (dB) at the frequency  $f$  (Hz),

$$G_1(f) = k_1 + 20 \log_{10} \left[ \frac{b_1^2 f^2}{(a_1^2 + f^2)(b_1^2 + f^2)} \right], \quad (11)$$

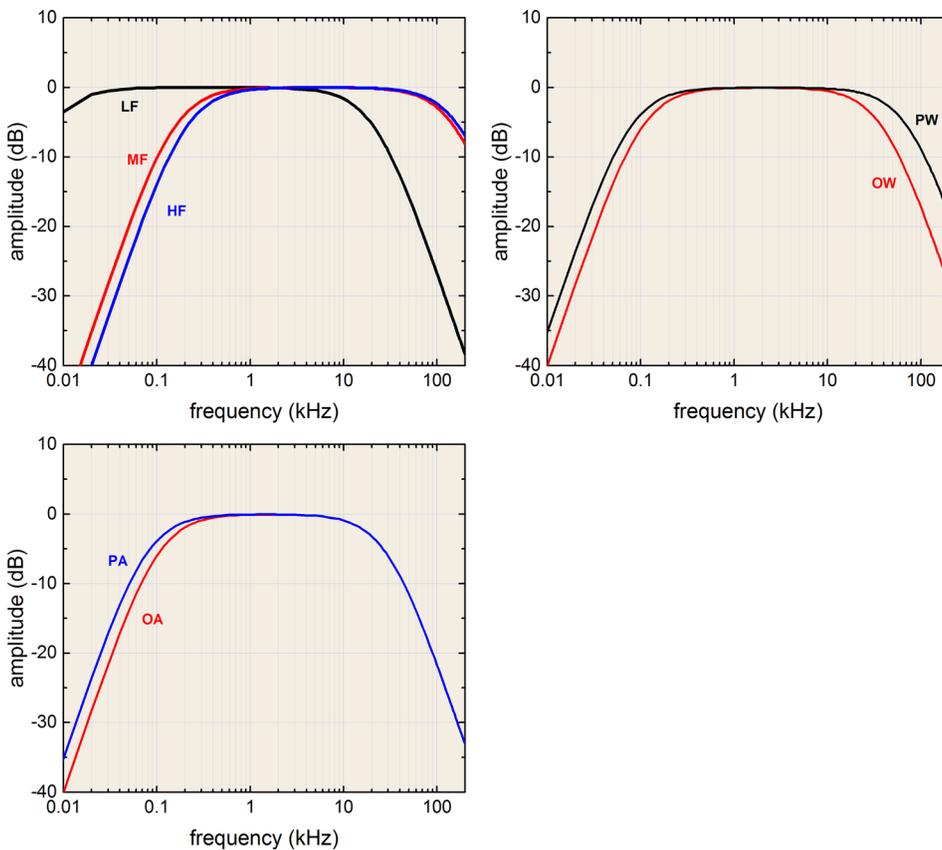


FIG. 18. (Color online) Navy type I weighting functions used for TAP phase 2 analyses (Finneran and Jenkins, 2012). LF—low-frequency cetaceans; MF—mid-frequency cetaceans; HF—high-frequency cetaceans; PW—phocids (in water), sirenians; PA—phocids (in air); OW—otariids, odobenids, mustelids, ursids (in water); OA—otariids, odobenids, mustelids (in air).

TABLE V. Marine mammal type II weighting function parameters for use in Eqs. (10)–(12). Parameters for the MF group were based on the 90-dB equal loudness contour for dolphins. Values for the LF and HF group were extrapolated from the MF group on a logarithmic basis relative to the hearing range for each group (Finneran and Jenkins, 2012).

Species group	$k_1$ (dB)	$a_1$ (Hz)	$b_1$ (Hz)	$k_2$ (dB)	$a_2$ (Hz)	$b_2$ (Hz)	Inflection point (Hz)
LF—Low-frequency cetaceans	-16.5	7	22 000	0.9	674	12 130	267
MF—Mid-frequency cetaceans	-16.5	150	160 000	1.4	7829	95 520	3000
HF—High-frequency cetaceans	-19.4	200	180 000	1.4	9480	108 820	3000

$$G_2(f) = k_2 + 20 \log_{10} \left[ \frac{b_2^2 f^2}{(a_2^2 + f^2)(b_2^2 + f^2)} \right], \quad (12)$$

the parameters  $a_1$ ,  $b_1$ , and  $k_1$  defined the type I component of the function and  $a_2$ ,  $b_2$ , and  $k_2$  defined the equal loudness-based component of the function (Table V).

For the Navy TAP phase 2 analyses, weighted thresholds for TTS were based on available TTS onset values obtained from representative species of mid-frequency cetaceans, high-frequency cetaceans, and pinnipeds. The shape of the type II weighting functions are shown in Fig. 20, the TTS exposures functions are shown in Fig. 21, and the corresponding weighted thresholds are provided in Table VI. The data obtained from mid-frequency and high-frequency cetaceans and pinnipeds were then extrapolated to the other species groups (for which no similar data were available).

For non-impulsive sounds, the mid-frequency cetacean threshold was based on the 195 dB re  $1 \mu\text{Pa}^2 \text{ s}$  SEL for TTS onset for dolphins at 3 kHz (Schlundt *et al.*, 2000). Since the type II weighting function amplitude at 3 kHz was -16.5 dB, the weighted TTS threshold became 178 dB re  $1 \mu\text{Pa}^2 \text{ s}$ . This same value was used for the low-frequency cetaceans, making the low-frequency and mid-frequency cetaceans equally susceptible to noise at the peak frequencies of their respective weighting functions. The high-frequency cetacean SEL threshold was derived from the impulse-noise TTS onset data for harbor porpoises published by Lucke *et al.* (2009) using the methods described by Southall *et al.* (2007). The weighted TTS thresholds for phocids in water and sirenians

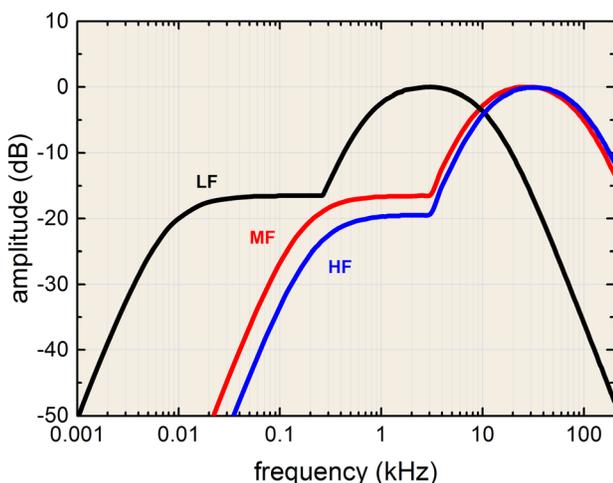


FIG. 20. (Color online) Navy type II weighting functions used for TAP phase 2 analyses (Finneran and Jenkins, 2012). LF—low-frequency cetaceans; MF—mid-frequency cetaceans; HF—high-frequency cetaceans.

were based on the harbor seal data reported by Kastak *et al.* (2005b). Thresholds for otariids, odobenids, mustelids, and ursids in water were based on the California sea lion data reported by Kastak *et al.* (2005b). TTS thresholds for phocids exposed to acoustic sources in air were based on data reported by Kastak *et al.* (2004) for a harbor seal. Thresholds for otariids, odobenids, and mustelids in air were based on California sea lion data reported by Kastak *et al.* (2004, 2007). PTS thresholds for non-impulsive sources were estimated by adding 20 dB to the onset TTS exposure for cetaceans and 14 dB for the other species groups (Finneran and Jenkins, 2012).

For impulsive sounds, SEL-based TTS thresholds for the low- and mid-frequency cetaceans were based on TTS data from a beluga exposed to water gun impulses (Finneran *et al.*, 2002). The TTS onset threshold for high-frequency cetaceans was based on TTS data from a harbor porpoise exposed to an underwater impulse produced from a seismic air gun (Lucke *et al.*, 2009). Thresholds for predicting TTS in phocid seals and sirenians were based on underwater TTS data from a harbor seal exposed to octave band noise (Kastak *et al.*, 2005b) and the extrapolation procedures described by (Southall *et al.*, 2007). Thresholds for otariids, odobenids, mustelids, and ursids in water were based on underwater TTS data from a California sea lion exposed to octave band noise (Kastak *et al.*, 2005b) and the extrapolation procedures described by Southall *et al.* (2007). The TTS thresholds for phocids, otariids, odobenids, and mustelids in air were based on unpublished data for harbor seals exposed to impulse noise (Southall *et al.*, 2007). PTS thresholds for impulsive sources for the TAP phase 2 analyses were estimated by adding 15 dB to the onset TTS thresholds (Finneran and Jenkins, 2012).

The type II functions were intended to improve upon the M-weighting/type I functions by accounting for the increased susceptibility to noise seen in the bottlenose dolphin TTS data at frequencies above 3 kHz. Figure 22 directly compares the type II functions to the M-Weightings from Southall *et al.* (2007). The downward shift in the HF type II exposure function relative to functions reported in Southall *et al.* (2007) reflects the addition of new data for HF cetaceans. The equal loudness weighting functions were not used by themselves because of the uncertainty regarding the weighting function amplitude at low frequencies below the range of the existing TTS and equal loudness data, i.e., the type I function was used at lower frequencies since there were no TTS or equal loudness data below 2.5 to 3 kHz. The type II weighting function represented a way to incorporate new data showing increased susceptibility to noise at higher

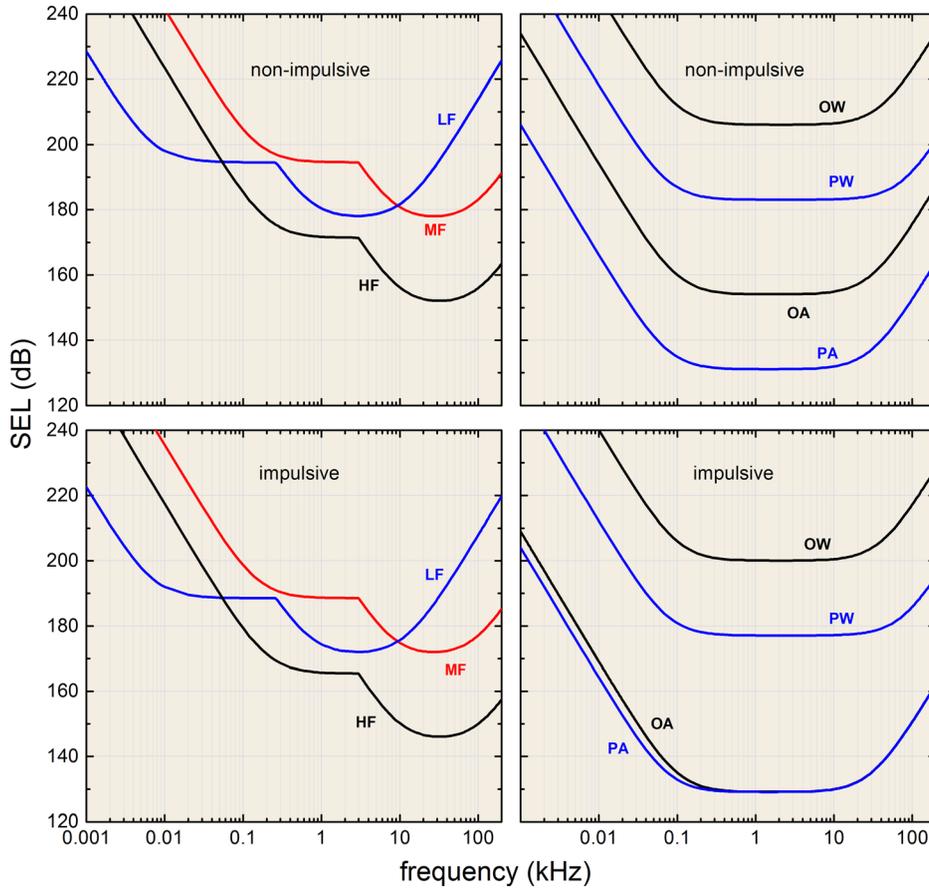


FIG. 21. (Color online) TTS exposure functions for U.S. Navy TAP phase 2 analyses (Finneran and Jenkins, 2012). LF—low-frequency cetaceans; MF—mid-frequency cetaceans; HF—high-frequency cetaceans; PW—phocids (in water), sirenians; PA—phocids (in air); OW—otariids, odobenids, mustelids, ursids (in water); OA—otariids, odobenids, mustelids (in air). Units for SEL are dB re  $1 \mu\text{Pa}^2 \text{ s}$  underwater (groups LF, MF, HF, PW, OW) and dB re  $(20 \mu\text{Pa})^2 \text{ s}$  in air (groups PA, OA).

frequencies with the broad weighting functions proposed by Southall *et al.* (2007).

### F. Porpoise equal latency functions (2014)

While multiple researchers have described equal latency functions for marine mammals (see Sec. IVC), Wensveen *et al.* (2014) used these functions to derive frequency weighting functions. More specifically, Wensveen *et al.* (2014) measured simple reaction time in a harbor porpoise, with tones at frequencies from 0.5 to 125 kHz and SPLs from 59 to 168 dB re  $1 \mu\text{Pa}$ , and used these data to make

inferences about equal loudness. The reaction time data were fit with exponential functions and used to create equal latency contours corresponding to reaction time values of 150, 160, 170, 180, 190, and 200 ms. The equal latency contours roughly paralleled the hearing threshold at relatively low levels and a flattening occurred with decreasing reaction time for frequencies  $<16$  kHz. Large changes in threshold occurred for relatively small changes in frequency in the region near best sensitivity ( $\sim 63$  to 125 kHz). The latency contours were therefore smoothed using the porpoise audiogram as a template. The smoothed functions were then inverted and normalized to yield a family of weighting

TABLE VI. Weighted SEL thresholds for TTS and PTS used for U.S. Navy TAP phase 2 (Finneran and Jenkins, 2012), shown by species grouping.

Species group	Non-impulse SEL threshold [dB re $1 \mu\text{Pa}^2 \text{ s}$ or dB re $(20 \mu\text{Pa})^2 \text{ s}$ ]		Impulse SEL threshold [dB re $1 \mu\text{Pa}^2 \text{ s}$ or dB re $(20 \mu\text{Pa})^2 \text{ s}$ ]	
	TTS	PTS	TTS	PTS
LF—Low-frequency cetaceans (type II WF)	178	198	172	187
MF—Mid-frequency cetaceans (type II WF)	178	198	172	187
HF—High-frequency cetaceans (type II WF)	152	172	146	161
PW—Phocids (in water), sirenians (type I WF)	183	197	177	192
PA—Phocids (in air) (type I WF)	131	145	129	144
OW—Otariids, odobenids, mustelids, ursids (in water) (type I WF)	206	220	200	215
OA—Otariids, odobenids, mustelids (in air) (type I WF)	154	168	129	144

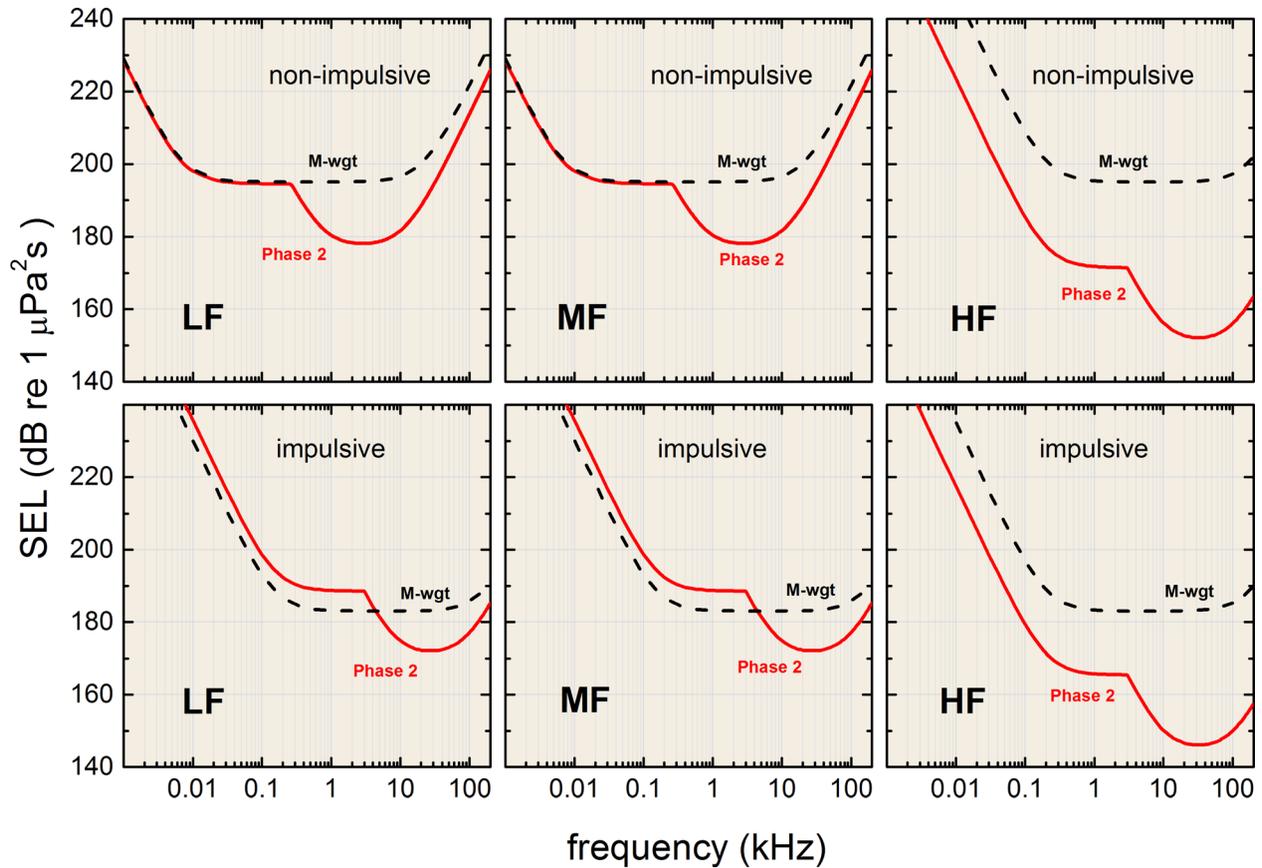


FIG. 22. (Color online) Comparison of cetacean TTS exposure functions for U.S. Navy TAP Phase 2 analyses (Finneran and Jenkins, 2012) (solid, red lines) and Southall *et al.* (2007) (black, dashed lines). Upper panels show exposure functions for non-impulsive noise, lower panels show exposure functions for impulsive noise. LF—low-frequency cetaceans; MF—mid-frequency cetaceans; HF—high-frequency cetaceans.

functions, as shown in Fig. 23. The weighting functions were essentially flat at high frequencies (above  $\sim 10$  kHz) and were linear with the logarithm of frequency at low frequencies. The slope of the linear-log curves at low frequencies ranged from 10 to 16 dB/octave, increasing with increasing reaction time (decreasing SPL).

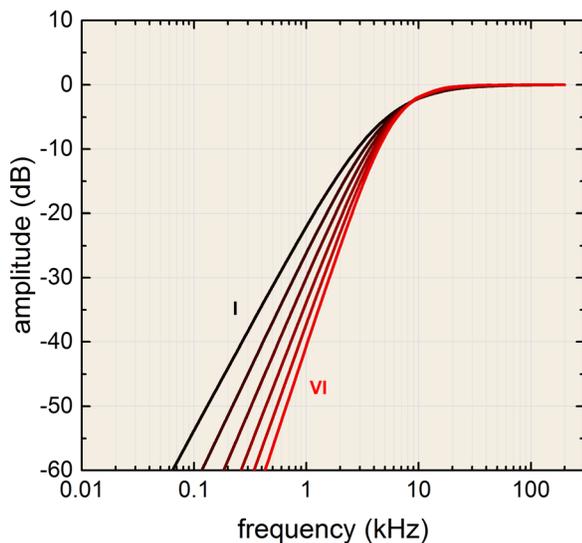


FIG. 23. (Color online) Auditory weighting functions derived from harbor porpoise equal latency data (Wensveen *et al.*, 2014). Source data are shown in Fig. 10.

### G. National Marine Fisheries Service (NMFS) weighting functions (2016)

In December 2013, the U.S. NMFS published initial draft guidance for assessing the effects of anthropogenic sound on marine mammal species under their jurisdiction (National Oceanic and Atmospheric Administration, 2013). The draft guidance identified received sound threshold levels above which individual marine mammals were predicted to experience NIHL for a wide range of underwater sound sources. Between 2013 and 2016, the draft guidance was modified to take into account new data, peer-review comments, public comments, and informal input from other U.S. governmental agencies. In August 2016, NMFS released its final Technical Guidance for Assessing the Effects of Anthropogenic Sound on Marine Mammal Hearing (National Marine Fisheries Service, 2016).

For the final Technical Guidance, NMFS adopted the same general approach proposed by the U.S. Navy for TAP phase 3 auditory weighting functions and TTS/PTS thresholds (Finneran, 2016). Essentially, weighting functions were modified from those used during the Navy phase 2 analysis (Finneran and Jenkins, 2012) to account for new data regarding marine mammal hearing sensitivity (e.g., Ghaul and Reichmuth, 2014; Ketten, 2014; Ketten and Mountain, 2014; Sills *et al.*, 2014; Cranford and Krysl, 2015; Sills *et al.*, 2015), equal latency contours (e.g., Reichmuth, 2013; Wensveen *et al.*, 2014; Mulsow *et al.*, 2015), and TTS (e.g.,

Kastelein *et al.*, 2012a; Kastelein *et al.*, 2012b; Finneran and Schlundt, 2013; Kastelein *et al.*, 2013a; Kastelein *et al.*, 2013b; Popov *et al.*, 2013; Kastelein *et al.*, 2014a; Kastelein *et al.*, 2014b; Popov *et al.*, 2014; Finneran *et al.*, 2015; Kastelein *et al.*, 2015a; Kastelein *et al.*, 2015b; Popov *et al.*, 2015).

Marine mammal species for which NMFS has regulatory oversight were divided into five groups for analysis: low-frequency cetaceans (LF), mid-frequency cetaceans (MF), high-frequency cetaceans (HF), phocids in water (PW), and otariids in water (OW). For each group, a unique weighting function was specified based on a generic band-pass filter described by

$$W(f) = C + 10 \log_{10} \left\{ \frac{(f/f_1)^{2a}}{\left[1 + (f/f_1)^2\right]^a \left[1 + (f/f_2)^2\right]^b} \right\}, \quad (13)$$

where  $W(f)$  is the weighting function amplitude (in dB) at the frequency  $f$  (in kHz). The shape of the filter is defined by the parameters  $C, f_1, f_2, a$ , and  $b$ . If  $a=2$  and  $b=2$ . Equation (13) is equivalent in form to the functions used to define Navy phase 2 type I and equal latency weighting functions and M-weighting functions (Southall *et al.*, 2007; Finneran and Jenkins, 2012), and the human C-weighting function (ANSI S1.42-2001). In addition to the weighting functions, exposure functions were defined by

$$E(f) = K - 10 \log_{10} \left\{ \frac{(f/f_1)^{2a}}{\left[1 + (f/f_1)^2\right]^a \left[1 + (f/f_2)^2\right]^b} \right\}, \quad (14)$$

where  $E(f)$  is the acoustic exposure as a function of frequency  $f$ , the parameters  $f_1, f_2, a$ , and  $b$  are identical to those in Eq. (13), and  $K$  is a constant adjusted to set the minimum value of  $E(f)$  to match the weighted threshold for the onset of TTS or PTS. The function described by Eq. (14) therefore reveals the manner in which the exposure necessary to cause TTS or PTS varies with frequency and allows the frequency-weighted threshold values to be directly compared to TTS data.

The parameters  $K, C, f_1, f_2, a$ , and  $b$  for each species group were obtained by first creating a representative,

composite audiogram for that group based on the available psychophysical hearing thresholds for species within the group. For mysticetes, psychophysical threshold data do not exist, therefore a representative audiogram was estimated using the (limited) available anatomical data and extrapolations from the other marine mammal composite audiograms (see Finneran, 2016). The exponent  $a$ , in Eqs. (12) and (13), was defined using the smaller of the low-frequency slope from the composite audiogram or the low-frequency slope of the equal latency contours (if available). The exponent  $b$  was set equal to two. The frequencies  $f_1$  and  $f_2$  were defined as the frequencies at which the composite threshold values were  $\Delta T$ -dB above the lowest threshold value. The value of  $\Delta T$  was chosen to minimize the mean-squared error between Eq. (14) and the non-impulsive, behavioral TTS data for the mid- and high-frequency cetacean groups. For species groups for which TTS onset data existed,  $K$  was adjusted to minimize the squared error between Eq. (14) and the steady-state (non-impulsive) TTS onset data. For other species,  $K$  was defined to provide the best estimate for the TTS onset at a representative frequency. The minimum value of the TTS exposure function was then defined as the weighted TTS threshold. Finally, the constant  $C$  was defined to set the peak amplitude of the function defined by Eq. (13) to zero.

The NMFS Technical Guidance weighting functions are depicted in Fig. 24. Weighting function parameters and acoustic exposure thresholds for non-impulsive and impulsive sounds are provided in Table VII. Figure 25 shows the TTS exposure functions that result when the weighting function amplitudes are applied to the TTS threshold values.

## H. Application of weighting functions to behavioral response thresholds

Although weighting functions are increasingly used to predict the auditory effects of noise exposures in marine mammals, they have not been consistently applied for predicting behavioral effects due to noise exposure, as in humans (e.g., annoyance). Animals cannot react to noise that is inaudible unless they are responding to either vestibular or somatosensory cues, so audible frequency range and hearing thresholds in noise must be taken into account when predicting behavioral effects. However, there is little evidence that the probability of behavioral disturbance or the severity of a response scales in a simple manner with the level of noise

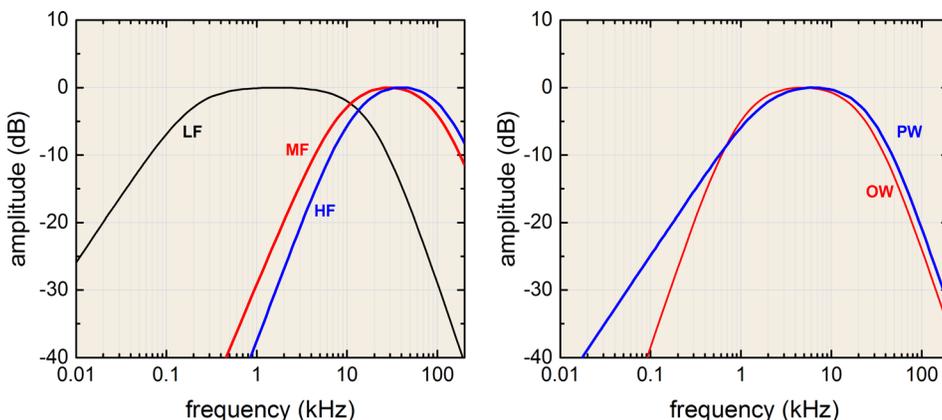


FIG. 24. (Color online) NMFS final Technical Guidance weighting functions for cetaceans, phocids, and otariids (National Marine Fisheries Service, 2016). Parameters required to generate the functions are provided in Table VII.

TABLE VII. Summary of weighting function parameters and TTS/PTS thresholds, in SEL with units of dB re  $1 \mu\text{Pa}^2 \text{ s}$ . Note that sirenians are not included as they are outside of NMFS jurisdiction.

Group	$a$	$b$	$f_1$ (kHz)	$f_2$ (kHz)	$C$ (dB)	Non-impulsive		Impulsive	
						TTS threshold	PTS threshold	TTS threshold	PTS threshold
LF	1	2	0.20	19	0.13	179	199	168	183
MF	1.6	2	8.8	110	1.20	178	198	170	185
HF	1.8	2	12	140	1.36	153	173	140	155
OW	2	2	0.94	25	0.64	199	219	188	203
PW	1	2	1.9	30	0.75	181	201	170	185

once it is above the animal’s threshold. At higher (supra-threshold) levels, limited evidence suggests a received level above which the amplitude of the signal does correlate with adverse behavioral reactions [e.g., see Houser *et al.* (2013a; 2013b) for examples of abandoned trained behaviors, avoidance, and the onset of what are believed to be annoyance behaviors]. Below this point, however, there is considerable variability likely driven by the context of the exposure (e.g., an animal’s prior experience with the sound source, behavioral state at the time of exposure, age, etc.). For example, a number of behavioral response studies have provided important information about animal responsiveness to acoustic exposure, but many questions about the influence of species tolerances, source proximity, prior experience with a sound

source, behavioral state during exposure (e.g., foraging), and social context on responsiveness remain (e.g., Sivle *et al.*, 2012; DeRuiter *et al.*, 2013; Goldbogen *et al.*, 2013; Antunes *et al.*, 2014; Miller *et al.*, 2014). Whether the ability to predict the behavioral responses of marine mammals to particular types and levels of noise can be improved through a weighting function approach, which takes differential sensitivity as a function of sound frequency into account, awaits further investigation and analyses.

### I. Remarks on marine mammal weighting functions

The development of marine mammal weighting functions has been heavily influenced by the need to quantitatively predict

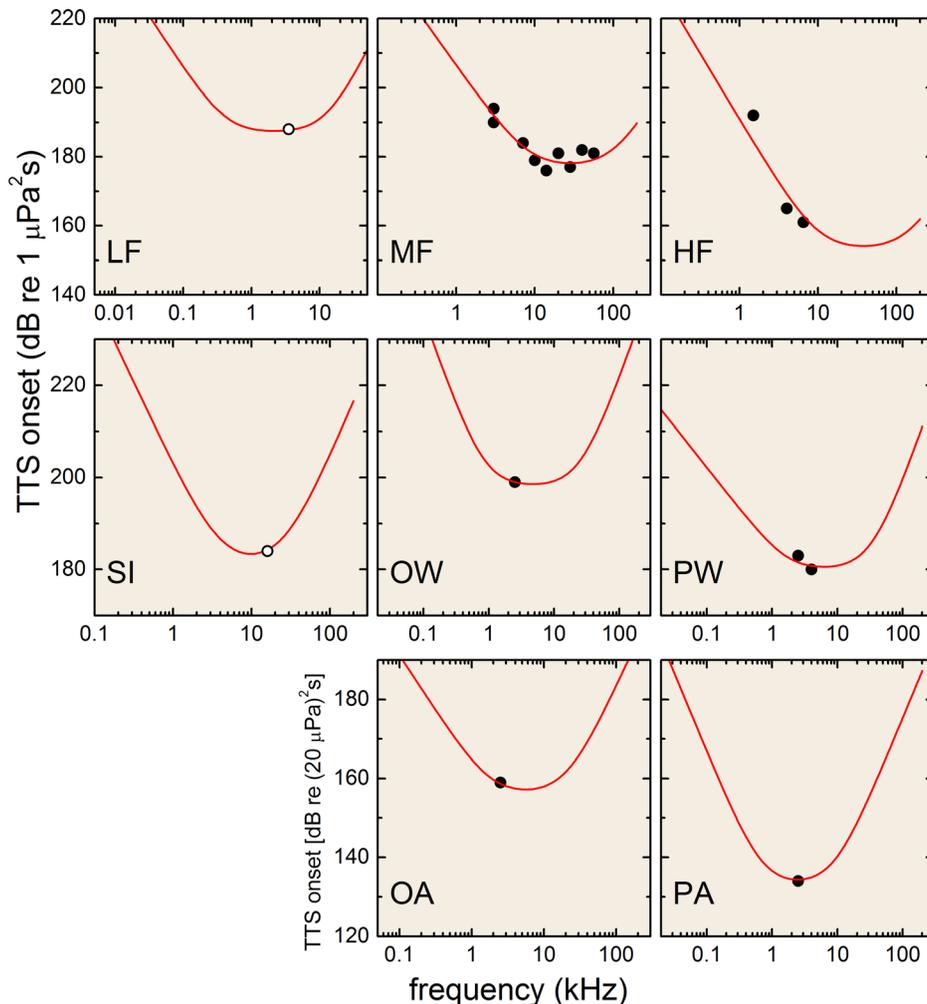


FIG. 25. (Color online) TTS Exposure functions resulting from NMFS final Technical Guidance weighting functions combined with TTS threshold levels (National Marine Fisheries Service, 2016). Filled symbols—onset TTS exposure data (in dB SEL) used to define exposure function shape and vertical position. Open symbols—estimated TTS onset for species for which no TTS data exist.

TTS/PTS (primarily driven by the U.S. MMPA), the large number of species for which functions are needed, and the limited amount of available data. Early efforts used simple methodologies based on very limited data; however, as more data have become available, the weighting functions have become increasingly complex. This has been primarily a result of the desire to follow a data-driven process (rather than one driven by expert opinion) and the need to extrapolate to species for which little or no data exist. It is likely that as more data become available, broad trends in the data will be revealed, and simpler approaches may be found acceptable. For example, as more data are collected showing the frequency dependency of TTS in various species, it may be found that an audiogram-based approach is a simple yet reasonably accurate alternative to complex extrapolation procedures involving curve-fits to TTS data.

It is likely that technical constraints on experimentation will continue to require that assumptions be made with certain aspects of marine mammal weighting function development. For example, limitations on the ability to produce sufficiently high levels of underwater sound at low frequencies, where the transmission voltage response of commonly used underwater transducers is poor, will make comparisons of signals with widely different loudness across the low frequency range difficult. This, in turn, will constrain the range of levels over which equal loudness contours could be obtained, thus requiring assumptions about the behavior of contours at higher levels of equal loudness. Similar issues exist with the obtaining of NIHL data at low frequencies. Where such limitations exist, and data cannot be obtained, assumptions will have to be made, possibly with a relatively weak scientific basis.

The application of marine mammal weighting functions is very different from that of human applications. The latter typically focuses on long duration exposures with continuously occurring noise (e.g., workplace exposures, annoyance from commercial activity). The former typically addresses shorter duration events with intermittent signals. These differences in application to some extent limit the degree to which the development of marine mammal weighting functions can rely on and be informed by human weighting functions. Indeed, even within the history of marine mammal weighting function development, it is uncertain how broadly applicable weighting functions are given the generally sparse data on which they are based and related assumptions about underlying auditory processes. For example, marine mammal weighting functions can be applied to weight the exposure to a single impulsive signal in order to predict the onset of NIHL. This is neither a common application in the use of human auditory weighting functions, nor are the processes of NIHL resulting from non-impulsive and impulsive noise sources necessarily the same (e.g., metabolic fatigue vs mechanical damage to the middle ear); yet, the same weighting is currently applicable to both non-impulsive and impulsive signals. How marine mammal weighting functions are applied should therefore continue to be periodically reviewed in order to consider new data that inform underlying assumptions related to their development and application.

## VI. GENERAL DISCUSSION AND RESEARCH RECOMMENDATIONS

The history and development of auditory weighting functions has progressed along somewhat related but separate pathways in humans and marine mammals. Each pathway is distinct with respect to the rationale for weighting function development, the process through which weighting functions have been developed, the time scales over which data supporting them were collected, and how the resultant weighting functions have been applied. Each also suffers from some assumptions that should be validated, and it remains uncertain in many cases as to whether the weighting functions most commonly used are ideal for the purposes to which they are applied. Nevertheless, at least in humans, the long history of A-weighting use across a broad spectrum of acoustic effects, along with little historical evidence of harm resulting from its application, suggests that there is utility in the application of A-weighting to effects ranging from annoyance to damage risk criteria. Because the use of auditory weighting functions in marine mammals has only been applied to regulation in the context of NIHL, it is uncertain whether there is broader utility given their particular basis and form. Furthermore, the utility of weighting functions for predicting NIHL in most marine mammal species is still in its infancy. Nevertheless, broad acknowledgement of the potential utility of auditory weighting functions for purposes of addressing marine mammal noise impacts exists, as exemplified in the inclusion of weighting functions within an ISO draft international standard for underwater acoustics terminology (ISO/DIS 18405.2, *Underwater Acoustics—Terminology*).

Auditory weighting functions for humans were originally developed to address issues of speech intelligibility across phone lines. The eventual utilization of equal loudness data as the basis for auditory weighting functions in humans is therefore intuitive given the original intent of their application. For marine mammals, weighting functions were initially developed to predict the incidence of hearing loss (TTS and PTS) following sound exposures. Similar to the rationale applied in the development of human weighting functions, it seems that the most appropriate data from which to derive weighting functions for marine mammals would therefore consist of exposure levels sufficient to cause TTS as a function of exposure frequency. If the objectives of applying the weighting function were different, it might be that other data types would be better suited to form the basis of the auditory weighting function. Indeed, it cannot be stated with certainty that the broad use of a single weighting function in humans is truly appropriate for the individual tasks to which it is applied. Furthermore, as observed from studies of NIHL in humans, a litany of complicating factors in predicting TTS or PTS (see Sec. III) suggests that there is perhaps no optimal weighting function to predict NIHL. For humans, the use of A-weighting, C-weighting, or the inverse function derived from the average human audiogram may produce nearly equal overall estimates of NIHL, depending on the noise characteristics. This is unlikely to be the case for marine mammals, which can have very different hearing

ranges and sensitivities across species, in addition to significant individual variability in NIHL susceptibility that likely exists within a species (similar to humans).

The development of auditory weighting functions for humans is based upon an immense amount of auditory data collected from thousands of individuals over many decades. This has provided confidence in the characterization of “normal” human hearing, including the expected variability in specific psychophysical measures evaluated at specific frequencies. By comparison, the available information for marine mammals is limited for every aspect of the source data used for the formation of auditory weighting functions. Audiograms across marine mammal species are sparse, particularly when split along the lines of those collected using behavioral, psychophysical methods and those collected via neurophysiological (i.e., AEP) methods. Sufficient source data for determining normal hearing (with appropriate means and variances) possibly exists for only one species of marine mammal, the bottlenose dolphin, and these data are primarily based on AEP measures above ~10 kHz. Similarly, data on NIHL only exist for a handful of individual marine mammals representing a small number of species, with TTS onset and growth data limited to a few frequencies. These constraints are complicated by the fact that—unlike humans in which a large amount of data exists for a single species—marine mammals represent more than a hundred species from several different taxonomic groups. As a result, the number of assumptions (e.g., broadly assuming hearing characteristics of untested species based on measurements in a few surrogate species) made in cross-species extrapolations are compounded. In short, despite significant advances in the understanding of marine mammal hearing over the last several decades, the paucity of species-specific data and small sample sizes remain an obvious and serious concern for the development of auditory weighting functions.

The human A-weighting function has been broadly applied to both physiological (e.g., NIHL) and behavioral (e.g., annoyance) effects associated with noise exposure whereas use of alternative, arguably more appropriate, weighting functions (e.g., C-weighting) have largely fallen into disuse. The scientific and policy decisions that led to the broad contemporary use of A-weighting are not clear. As noted by [McMinn \(2013\)](#), the decision to more broadly rely on A-weighting as a standard feature of sound level meters and other acoustic recording devices may have occurred several decades ago through agreement within ISO. At a minimum, such broad adoption within the realm of measurement instrumentation likely influenced later policy decisions related to measuring acoustic exposures in humans. Nevertheless, there appears to be little scientific justification or support to indicate that A-weighting is the optimal weighting function to be used across the wide range of applications for which it is currently applied. It may well be that alternative weighting functions are better suited to different situations. For example, C-weighting is potentially more suitable for assessing physiological impacts resulting from exposure to high level noise than A-weighting since a flattening of auditory filters with progressively higher received levels occurs. Similar arguments exist for marine mammals, which have had different auditory weighting functions developed to

account for variation in the frequency range of hearing and sensitivity across species groupings. It remains unknown as to whether and to what degree the various marine mammal weighting functions can be suitably applied to effects other than NIHL, such as noise-induced behavioral disturbance and masking.

Current marine mammal auditory weighting functions would greatly benefit from research addressing specific aspects of their underlying assumptions, data limitations, and the manner in which they are applied. Below, a number of research recommendations are offered, in no particular order, for consideration that would increase the robustness of marine mammal auditory weighting functions and help to validate their application. It is important to note that these recommendations are not accompanied by recommendations for policy changes or regulatory development, nor recommendations for applying weighting functions in the estimation of acoustic impact on marine mammals resulting from noise exposure. Such recommendations fall outside the scope of this review, which is directed at examining the science underlying the development of weighting functions.

### **A. Increase effort on audiogram acquisition**

Determining normal hearing in any species, and defining it both with means and variances, requires obtaining audiograms on numerous animals that are free of pathological and age-related otological issues. Confidence in the ability to apply weighting functions to issues of human concern is rooted in extensive data on the hearing abilities of normal, otologically healthy young persons. In contrast, audiograms do not exist for most marine mammal species, and for those species with existing data, an insufficient number of audiograms exist for characterizing “normal” hearing for any species. There are a few species measured under laboratory conditions by which sufficient sample sizes could be obtained (e.g., bottlenose dolphins, California sea lions), albeit at a much smaller scale than has been completed in humans. Wherever possible, high-quality behavioral audiograms should be obtained with species for which few auditory data are available. This should also be done for species that have some available auditory information so that sample sizes can be expanded. In addition, sample size can be increased by AEP testing of marine mammals. For example, short-term access to some pinnipeds can be provided through rehabilitation facilities, and for less accessible species, through targeted capture-test-release efforts, especially for those species that congregate on rookeries. This could provide population level measures of hearing for pinnipeds via AEP methods, although such methods would benefit from further refinement because of limited application to certain species (e.g., phocid seals). For odontocete cetaceans, data collection will likely be obtained slowly and through the testing of stranded individuals. Since opportunities for testing live-stranded and rehabilitating odontocetes are few, concerted efforts to test stranded species will need to be supported by the agencies that oversee their disposition. However, health assessments of small odontocetes in which animals are temporarily captured and held might also

provide opportunities to expedite data collection for some species (e.g., bottlenose dolphins, belugas). A prime goal of evaluating audiogram data should be the assignment of novel species into valid functional hearing groupings (e.g., mid-frequency cetacean), as well as consideration of the extent to which other hearing data (e.g., TTS, masking) from commonly tested species can be extrapolated to exotic species.

The determination of hearing in mysticete cetaceans is a much larger problem and is given its own discussion in Sec. VIF.

### **B. Investigate the relationship between AEP measures and behavioral measures of auditory perception**

Although behavioral studies of marine mammal hearing remain the accepted standard, the use of AEP methods to test various aspects of marine mammal hearing have rapidly proliferated. The two approaches, however, do not provide the same results; behavioral measures provide an integrated “whole” animal response, including perceptual processing, learning and decision making by the animal, whereas AEPs are a measure of neural activity arising from specific sites within the brain (e.g., brainstem vs cortical origins). It is likely that AEPs will continue to be used to study marine mammal hearing and that audiometric data obtained with AEPs will eventually gain greater informative power in the generation of auditory weighting functions. For this to occur, studies need to be undertaken that establish the relationship between AEP and behavioral measures of hearing. Thus far, a single study has reported behavioral and evoked potential measures collected simultaneously with a marine mammal (Schlundt *et al.*, 2007). Future studies should directly compare measures of TTS and hearing obtained behaviorally and via AEP, as has been reported for bottlenose dolphins (Finneran *et al.*, 2007; Finneran *et al.*, 2015) and the false killer whale (*Pseudorca crassidens*, Yuen *et al.*, 2005). This would allow TTS studies involving only AEPs to be more adequately interpreted and applied. Similarly, differences in thresholds (and the variability in differences) obtained with AEP and behavioral methods should be better characterized as AEP methods are likely the means by which the hearing of more exotic species will be tested. Comparisons between behavioral measures of masking and AEP measures should be made to determine how well specific AEP masking effects correlate with behavioral measures of masking. In all cases, differences are expected between the two approaches, as they measure different auditory phenomena. Further, there are limits to the information that can be obtained via AEPs in the lower regions of an animal’s hearing range. Nevertheless, given the ability to apply AEP methods in field situations (e.g., stranded cetaceans), and thus the greater number of species that can be tested, it is likely that the role of AEP methods in providing data informative to auditory weighting function development will increase.

### **C. Understand data variability in studies of NIHL (increase sample sizes)**

Studies in humans have demonstrated that there is broad inter-individual variability in susceptibility to NIHL. An

understanding of the degree of individual variability, and the causes of this variability, have only been possible due to the large number of studies and subjects used in human NIHL research. Marine mammals will certainly show individual variability in susceptibility to NIHL, but current knowledge of the degree of variability for a given fatiguing stimulus is severely limited by the small number of subjects studied to date (see Finneran, 2015). Studies of NIHL in marine mammals should be replicated and sample sizes increased. Ideally, studies should look for a cross-sectional representation within the species tested (e.g., different ages, both genders). Studies of additional species—focusing on species of particular concern or those that may be poorly represented by available data due to body size or hearing range (e.g., killer whales, beaked whales, mysticete whales)—are, of course, also warranted. Furthermore, studies of NIHL to underwater and in-air sound exposures would be informative to understanding NIHL processes and sources of variability for amphibious species, such as the pinnipeds.

### **D. Acquire additional frequency-specific data on NIHL in marine mammals**

The frequency-specific nature of TTS-onset has a profound influence on the shape of U.S. Navy and NMFS marine mammal auditory weighting functions. However, the degree to which this frequency dependence informs the weighting function is not fully realized because there are a limited number of frequencies at which TTS has been measured, and these only in a few marine mammal species. Information from dolphins, belugas, and porpoises demonstrates that the region of best hearing sensitivity does not necessarily contain the frequencies at which the animal is most susceptible to NIHL [see Finneran (2015), for review]. Because of potential differences between NIHL susceptibility and the absolute hearing sensitivity for any given frequency, the audiogram does not appear to be the best predictor of NIHL. To adequately account for the frequency dependence of NIHL, sufficient testing of TTS across the range of hearing should be performed. This requires TTS growth curves across the audible frequency range for representative species in different hearing groups. From such growth curves, TTS onset and PTS onset could be quantified and estimated, respectively. These data are needed at least for a small number of individuals with normal hearing. To obtain these data will be both expensive and time consuming, but it will allow a more accurate characterization of how frequency-specific NIHL susceptibility affects the auditory weighting function and allow a detailed evaluation as to whether audiograms may provide acceptable approximations to TTS onset. Furthermore, weighting functions can *a priori* be used to make predictions of the occurrence of TTS, which can then be validated through the collection of NIHL data.

### **E. Improve comparability of audiometric information through standardization**

The high degree of individual variability observed in human studies of hearing, loudness, NIHL, etc., is likely to also be observed in marine mammals. One potential source

of variability that can be accounted for and controlled is that due to differences in experimental methods. As has been done for particular lines of human audiometric research, the contribution of measurement variability due to methodological differences can be minimized through standardization of research methods. Standards could be formally and collaboratively developed through existing infrastructure (e.g., ISO, ANSI, etc.). Further, should funding agencies determine to invest in a particular area of study (e.g., TTS resulting from tonal stimuli), they could convene investigator meetings in advance of commencing research to ensure that research methods between laboratories are as consistent as possible given the logistical limitations imposed by the study animals (e.g., dolphins vs sea lions). To ensure the greatest degree of comparability, such meetings could be convened before, during and following the conclusion of the research, or occur in a multi-institution framework.

#### **F. Mysticetes: Audiogram prediction, validation of auditory models and weighting functions**

The hearing capability of mysticete whales is arguably the largest unknown and greatest impediment to predicting how anthropogenic sound exposure might affect mysticetes. Anatomical models aimed at predicting mysticete hearing sensitivity exist (Houser *et al.*, 2001; Cranford and Krysl, 2015), but lack validation against any empirical measure of hearing and the error in the accuracy of their predictions cannot be quantified. The models used to date also do not simulate inner ear neural output, and are therefore incapable of providing well-informed estimates of absolute sensitivity at any frequency. Ideally, audiometric data should be obtained from mysticetes (most likely through AEP of stranded individuals or behavioral response studies of free-ranging individuals) that can be used to validate various aspects of anatomical model assumptions or predictions. Use of vocalization frequency ranges to predict hearing range is of limited utility in this regard, as vocalization frequencies do not strictly delineate the limit of hearing and often do not overlap with the most sensitive region of hearing (Heffner and Heffner, 1992; Mulsow and Reichmuth, 2010). In lieu of direct measures of hearing sensitivity, or until such time that it is possible, modeling approaches similar to that used for mysticetes could be created for odontocetes and then validated through behavioral or neurophysiological approaches. Many odontocete species are more readily available, either at zoological facilities or through stranding and rehabilitation, and are logistically more amenable to currently available research methods. Validation of odontocete anatomical models that are then expanded to mysticetes in the absence of such validation would increase confidence in the results obtained.

The current approach to developing marine mammal auditory weighting functions requires parameter values for the upper- and lower- cutoff frequencies. Although anatomy and anatomical models have been used to qualitatively or quantitatively predict hearing ranges in mysticetes (e.g., Parks *et al.*, 2007; Manoussaki *et al.*, 2008; Tubelli *et al.*, 2012), cutoff frequencies are currently only “best guesses”

for mysticetes based on trends extrapolated from other marine mammals, and need to be determined to provide confidence in predictions utilizing the weighting functions. Direct measures of hearing (e.g., AEP) in any mysticete species would provide a wealth of information relevant to validating model predictions, determining if extrapolations from other species are appropriate, and providing data that can directly be used in designing mysticete weighting functions. Alternative feasible approaches include directed behavioral response studies aimed at addressing basic questions about sound reception (e.g., determining conservative upper- and lower-frequency limits of hearing), and additional independent efforts to model auditory sensitivity using similar and novel approaches.

#### **G. Explore the extent to which data and models from terrestrial mammals (including humans) are applicable to marine mammals**

There are a number of approaches used to develop human weighting functions that have potential utility with respect to marine mammal auditory weighting functions. For example, equal loudness contours measured for humans can be derived from data on the function of the auditory periphery (e.g., the GM model). If the function of the auditory periphery in marine mammals were essentially similar to those of other mammals used to develop models of loudness, it could be possible to adapt or develop models to predict equal loudness contours for marine mammals without having to directly measure loudness (or a loudness proxy such as reaction time). In a similar manner, exploration of human/terrestrial models and comparisons of marine mammal auditory processes to terrestrial mammal auditory processes might shed light on where extrapolations from terrestrial studies would be most useful and appropriate in guiding the development of marine mammal auditory weighting functions. For example, it might be useful to more extensively explore the relationship between TTS and PTS (using perhaps AEP measures) in an animal model where PTS is not deemed a catastrophic experimental outcome, and where the upper-frequency limits of the animal model’s audiogram is closer to that of marine mammals (such as mice).

The use of reaction time, despite some limitations, is still a behavioral measure that might provide a valid indicator of a sound’s loudness. One way to study the validity of reaction time as a measure of loudness is to compare the results across animal species, including humans; data from humans would be used to indicate the exact relationship between loudness judgments (i.e., equal loudness judgments) and reaction times. In so doing, to the extent possible, the same stimuli and experimental procedures should be used for all measurements across methods and animals. More experiments like the one conducted by Finneran and Schlundt (2011), in which measures like discrimination and detection are used to determine how sounds of different frequency and level affect discrimination and detection performance, might be useful. Such experiments are time consuming, but just a few demonstrations could significantly improve the confidence in derived weighting functions.

## H. Conduct studies in marine mammals to determine if weighting functions developed as damage risk criteria for NIHL are appropriate for masking and other behavioral measures

As described throughout the review of human research, weighting functions have been applied to a number of effects aside from NIHL (e.g., performance and judgment, vigilance, anxiety and stress). They form the basis of numerous annoyance metrics and are applied in various local, state and federal laws and ordinances governing acceptable levels of noise. Federal laws within the United States exist that require the behavioral disturbance of marine mammals due to anthropogenic noise exposure be accounted for (Marine Mammal Protection Act of 1972, as amended in 1994). The NMFS, which oversees the regulation of marine mammals, regulates the potential for behavioral disturbance (harassment) with received SPLs of 160 and 120 dB for impulsive and continuous type sounds, respectively (see 70 Fed. Reg. 1871). However, as of the writing of this review, the behavioral disturbance of marine mammals due to exposure to U.S. Navy sound sources is determined through a dose-response relationship where the maximum unweighted SPL received by a marine mammal over a period of sound exposure determines the probability that a behavioral disturbance will occur (e.g., Department of the Navy, 2013). Debates exist as to whether this approach is appropriate (e.g., what other factors, such as source proximity, should be considered). Whether the received sound should be weighted as part of the disturbance determination has been one of the topics of debate, but it has never been assessed as to whether weighting the received SPL (or other sound measure) improves model predictions for disturbance. Similarly, frequency-dependent weighting has proved useful in understanding masking and saliency of linguistic signals in humans, but its application to masking and signal saliency in the regulation of marine mammal noise issues has been entirely overlooked. Investigations of whether weighting functions can improve predictions of impact to marine mammal behavior should be completed. This could be achieved by conducting behavioral response studies in species for which weighting functions are best defined, or could be achieved retrospectively with current weighting functions applied to behavioral response studies already completed. Similarly, studies should be conducted to determine the applicability of weighting functions to predict when normal hearing is constrained (masked) by elevated noise levels. In the case of masking, psychophysical studies could examine how weighting may relate to masking in individual animals, not only with respect to signal detection, but also with respect to signal saliency, localization, and discriminability.

### I. Determine how frequency-dependent auditory phenomena scale with acoustic signal strength

The physiological response of the auditory system is not static and changes in response to the level of sound it receives. This has consequences for auditory perception. For example, auditory filters broaden as the stimulus level increases, particularly along the low-frequency tail of the

filters (Lutfi and Patterson, 1984; Patterson and Moore, 1986; Moore and Glasberg, 1987). The level-dependent changes in auditory filter shape have been associated with nonlinear growth in masking (Lutfi and Patterson, 1984). In many such studies in the human literature, level dependent changes in performance are attributed to the non-linear compression that occurs within the inner ear, most likely as a result of outer hair cell function (see Bacon *et al.*, 2004). The implications of findings such as these are that auditory perception likely scales with acoustic signal strength (e.g., as observed in the level-dependence of loudness functions). Furthermore, research with a bottlenose dolphin, false killer whale, and beluga demonstrate an ability to vary hearing sensitivity in anticipation of moderate to high-level noise exposures (Nachtigall and Supin, 2013, 2014, 2015; Nachtigall *et al.*, 2016). These findings raise the question as to whether phenomenon studied at low levels of sound exposure (e.g., near threshold) are truly representative of auditory processes at moderate to high levels of sound exposure.

Investigations into the dynamic relationship between acoustic signal strength and features of the auditory system that affect perception should be conducted in marine mammals to better understand, characterize, and model level-dependent phenomena. Specifically, characterization of the impact of stimulus strength on auditory filter shape would better inform how auditory weighting functions might change with stimulus level. This, in turn, would allow better predictions of perceptual impacts to marine mammals exposed to noise across a range of stimulus levels. While the manner in which perceptual phenomena (such as loudness) scale with signal strength is difficult to study, other aspects of hearing (such as auditory filter shape) can be readily tested using available methods and may prove equally useful.

## VII. CONCLUSIONS

The frequency-dependent nature of hearing in humans and marine mammals has led to the natural adoption of weighting function approaches to address effects related to noise exposure (NIHL in marine mammals, NIHL and various perceptual and behavioral effects in humans). While a number of the steps necessary for the further development of weighting functions for marine mammals will likely parallel, to some degree, those already taken with humans (e.g., increases in sample sizes), many will reflect the unique circumstances presented by marine mammal species (e.g., difficulty in obtaining mysticete data, large number of species, indirect measurements of loudness). Along these lines, the following conclusions are presented from the comparative material covered in this review.

- (1) Auditory data for humans are extensive as they are limited to one easily accessible and testable species. Such sample sizes will not be achievable for any species of marine mammal, and future research should strive to determine general processes that can be extrapolated across species within defined functional hearing groups.
- (2) The A-weighting function developed for humans, based on the 40-phon equal loudness contour, arose out of the

specific need to determine speech intelligibility over phone lines. Nevertheless, A-weighting has seen continued application in numerous human regulatory contexts. It is possible that weighting functions for marine mammals therefore need not be strictly tied to a particular equal loudness contour in predicting a diverse set of noise effects. However, additional research on the utility of broader application is required [see item (4), below].

- (3) The refinement of auditory weighting functions for marine mammals has been primarily driven by the desire to predict TTS onset levels. The evolution of these weighting functions has resulted in an approach—based on adapting the audiogram to better fit the existing TTS data—that appears to predict the frequency dependence of TTS onset primarily for mid- and high-frequency cetaceans. Additional TTS data across frequencies—especially outside of the range of best hearing sensitivity—and species groupings will be necessary to further validate or refine current approaches.
- (4) Future research efforts should focus on producing weighting functions that provide accurate predictions of a particular physiological, perceptual, or behavioral effect, ideally in a manner that is robust for a wide variety of noise types.

## VIII. ACKNOWLEDGMENTS

The authors wish to thank the large number of reviewers that took the time to read and comment on prior versions of this manuscript, specifically A. Scholik-Schlomer, S. Labak, T. Brookens, R. Dekeling, C. Lemont, F.-P. Lam, M. Ainsle, S. von Benda-Beckman, R. Gisiner, R. Gentry, R. Kastelein, P. Wensveen, A. Ruser, and V. Reyes. This review was funded by the International Association of Oil and Gas Producers E&P Sound and Marine Life Joint Industry Programme.

## APPENDIX: DEFINITION OF TERMS

Where possible, the terms in this report are consistent with ANSI/ASA S1.1-2013, *Acoustic Terminology*. The standard definition of some terms contained in S1.1 does not always agree with the use of the term in the literature. In these cases, using the standard definition might hinder the ability for readers of the review to understand the literature cited in the report. This appendix is a combination of standard definitions and the author's definitions based on what they believe is common usage in the literature. When a standard definition is used, a reference to the number (X.XX) of the term in ANSI/ASA S1.1 or ANSI/ASA S3.20 will be indicated. This appendix is not intended to be exhaustive. Any term defined within the document is not re-defined here.

### 1. Definitions from ANSI/ASA S1.1-2013

2.06 peak sound pressure: Greatest absolute value of instantaneous sound pressure within a specified time interval. Unit, pascal (Pa).

Note that in this document, peak sound pressure is often converted to a decibel value and referred to as the peak sound pressure level [see Eq. (4)]. The sound waveform might or might not be filtered by a weighting function before the peak sound pressure is determined. It is assumed to be unweighted unless specified otherwise.

2.59 sound pressure: Total instantaneous pressure at a point in a medium minus the static pressure at that point. Unit, pascal (Pa); symbol,  $p$ .

2.70 sound intensity: Average rate of sound energy transmitted in a specified direction at a point through a unit area normal to this direction at the point considered. Unit, Watt per square meter ( $\text{W}/\text{m}^2$ ). Where sound energy (2.66) is the total energy in a given part of a medium minus the energy that would exist at that same part of the medium with no sound waves present. Unit, Joule (J).

3.11 spectrum level: Level of the limit, as the width of the frequency band approaches zero, of the quotient of a specified power-like quantity distributed within a frequency band, by the width of the band. The words "spectrum level" should be preceded by a descriptive modifier. Unit, decibel (dB).

11.05 loudness level: Of a sound, the median sound pressure level in a specified number of trials of a free progressive wave having a frequency of 1000 Hz that is judged equally loud as the unknown sound when presented to listeners with normal hearing who are facing the source. Unit, phon.

### 2. Definitions from ANSI/ASA S3.20-2015

C11.29 sensation level; level above threshold: For an individual listener and a specified sound signal, amount by which a sound pressure level or force level exceeds the hearing threshold for that sound. Unit, decibel (dB).

### 3. Deviation from standard definitions (or not defined in an ANSI standard)

Amplitude: This document defines amplitude as the magnitude of a transfer function or filter. That is, transfer functions (filters) will be expressed as amplitude in dB as a function of frequency (in Hz or kHz). In these cases, amplitude is a pressure-like term. Typically, the amplitude of the pass band is 0 dB and the amplitude is expressed as a negative value relative to 0 dB (i.e., as attenuation in dB).

dBA, B, C, D, Z, H, Ht, M: dB followed by any of the aforementioned letters (e.g., dBA) refers to a SPL measurement where the waveform has been modified by the indicated weighting scale (e.g., A-weighting). That is, the decibel level of each spectral component of the sound has been added to the amplitude (expressed as decibels) of the transfer function (filter) at each corresponding frequency representing the particular weighting scale (see Fig. 1).

Hearing level (HL): This document uses dB HL to indicate when the reference pressure for a SPL measured at a particular tonal frequency is a threshold of human hearing specified in a national or international standard at that particular tonal frequency, for a given earphone or speaker, and in the former, in a given coupler. That is, dB HL indicates the

difference in the SPL of the sound being measured and the SPL specified in a standard as a threshold of hearing. Further explanation of dB HL is provided in the document.

**Sound level** (often shortened to level): Sound level in this document generically refers to any decibel measure of sound magnitude, e.g., sound intensity, sound pressure level, hearing level, sensation level, etc. It is not used in the sense of ANSI/ASA S1.1-2013, note 3 of definition 3.09 sound pressure level, which narrowly defines it as “If a stated frequency weighting is applied to the sound-pressure signal, then the result is a sound level, not a sound pressure level.”

**Weighting:** The term “weighting” preceded by a letter, e.g., A-weighting, will refer the use of the specified weighting scale (i.e., the A scale) when a decibel measure of sound magnitude is made. The document will assume that the weighting measure is an SPL measurement unless otherwise stated. When a sound is measured by a sound level meter, it is frequency weighted by a “weighting scale,” and the sound is time integrated. Both weighting and time integration functions are specified in ANSI/ASA S1.4-2014. In most cases the weighting functions are the A, C, and Z (flat) weighting, as is explained in the document, and the time integration is either fast (F) or slow (S), as described in ANSI/ASA S1.4-2014. If other time integration functions are used, they are specified.

<sup>1</sup>Sound level, frequency, and spectrum are physical terms. Loudness, pitch, and timbre are perceptual terms. There is not a one-to-one correspondence between these physical and perceptual terms (e.g., loudness can vary when either frequency or level is varied).

<sup>2</sup>Soon after the establishment of the decibel, Riesz (1928) measured the discrimination thresholds for changes in sound level for tones of different frequencies and levels. The threshold for discriminating a change in sound level varied from approximately 0.5 dB to 2.5 dB, with 1 dB being the most common difference in threshold over a large range of intensities and frequencies (except at the extremes of audibility in terms of frequency and level). Several problems exist with the method used by Riesz (1928) to measure level discrimination. Studies (e.g., see Jesteadt *et al.*, 1977) have since measured level discrimination with different procedures and shown the just detectable difference in tonal level is approximately, but not exactly, equal to 1 dB over a considerable range of level and frequency.

<sup>3</sup>Starting in the 1880s, and more significantly by the mid-20th century, the CGS (centimeter, gram, second) system was gradually superseded for scientific purposes by the MKS (meter–kilogram–second) system. This, in turn, developed into the current SI standard. For mechanical systems, the differences among the various unit systems is straight forward, but is more complicated for electromagnetic phenomena.

<sup>4</sup>Several other studies at Bell Labs involving sound level preceded Sivian and White, e.g., Wegel and Lane (1924), but Sivian and White were the first to clearly lay out the procedures for measuring and calibrating sound level in the determination of the threshold of hearing.

<sup>5</sup>The procedures used to measure MAF and MAP thresholds both should theoretically reflect the sound pressure level at the eardrum required by the listener to obtain a threshold estimate of tonal detection. However, MAF and MAP thresholds of hearing differ on average by about 6 dB at frequencies <2 kHz, and less above 2 kHz. This problem was recognized by Sivian and White (1933) and they proposed several explanations for the difference. More recent explanations [see Yost and Killion (1997)] have argued [somewhat similar to Sivian and White (1933)] that most of the differences between MAF and MAP estimates of the thresholds of hearing can be explained by taking into account resonances in the outer ear canal.

<sup>6</sup>Interview with Dr. Harvey Fletcher by Vern Knudsen and W. J. King at the University of California, Los Angeles on May 15, 1964. The interview was recorded by the Center for History of Physics of the American Institute of Physics.

<sup>7</sup>For examples, see the Wyle primer related to aviation noise available and its effects at <http://www.rduaircraftnoise.com/rduaircraftnoise/noiseinfo/downloads/NoiseBasicsandEffects.pdf> (Last viewed 2/16/2017).

<sup>8</sup>From Allen (1996), “When several tones are far apart in frequency, their loudnesses (the output of the cochlea) add; but when they are very close together, the acoustic intensities (the input) add.” “Far apart” and “close together” are defined in terms of the critical bandwidth.

<sup>9</sup>Marine Mammal Protection Act of 1972, as amended in 2007, <http://www.nmfs.noaa.gov/pr/pdfs/laws/mmpa.pdf> (Last viewed 2/16/2017).

<sup>10</sup>TTS data for bottlenose dolphins were first reported by Ridgway *et al.* (1997) in a technical report. These data were later combined with additional data for belugas and dolphins and published by Schlundt *et al.* (2000).

Acoustical Society of America (1936). Z24.3-1936, “American tentative standards for sound level meters for measurement of noise and other sounds,” *J. Acoust. Soc. Am.* **8**, 147–152.

Allen, J. B. (1996). “Harvey Fletcher’s role in the creation of communication acoustics,” *J. Acoust. Soc. Am.* **99**, 1825–1839.

ASTM (2008). E1130-08, *Standard Test Method for Objective Measurement of Speech Privacy in Open Plan Spaces Using Articulation Index* (ASTM International, West Conshohocken, PA).

Antunes, R., Kvadshheim, P. H., Lam, F. P., Tyack, P. L., Thomas, L., Wensveen, P. J., and Miller, P. J. (2014). “High thresholds for avoidance of sonar by free-ranging long-finned pilot whales (*Globicephala melas*),” *Mar. Pollut. Bull.* **83**, 165–180.

Awbrey, F. T., Thomas, J. A., and Kastelein, R. A. (1988). “Low-frequency underwater hearing sensitivity in belugas, *Delphinapterus leucas*,” *J. Acoust. Soc. Am.* **84**, 2273–2275.

Babkoff, H., Pratt, H., and Kempinski, D. (1984). “Auditory brainstem evoked potential latency-intensity functions: A corrective algorithm,” *Hear. Res.* **16**, 243–249.

Bacon, S. P., Popper, A. N., and Fay, R. R. (2004). *Compression: From Cochlea to Cochlear Implant* (Springer, New York), p. 228.

Björk, E., Nevalainen, T., Hakumäki, M., and Voipio, H.-M. (2000). “R-weighting provides better estimation for rat hearing sensitivity,” *Lab. Anim.* **34**, 136–144.

Blakeslee, E. A., Hynson, K. N., Hamernik, R. P., and Henderson, D. (1978). “Asymptotic threshold shift in chinchillas exposed to impulse noise,” *J. Acoust. Soc. Am.* **63**, 876–882.

Broadbent, D. (1981). “Effects of moderate levels of noise on human performance,” in *Hearing Research and Theory*, edited by J. Tobias and E. Schubert (Academic Press, New York), pp. 165–186.

Buck, K. (1982). “Influence of different presentation patterns of a given noise dose on hearing in guinea-pig,” *Scand. Audiol. Suppl.* **16**, 83–87.

Burkard, R., and Don, M. (2015). “Introduction to auditory-evoked potentials,” in *Handbook of Clinical Audiology*, edited by J. Katz, M. Chasin, K. English, L. Hood, and K. Tillery (Wolters Kluwer Health, Philadelphia, PA), pp. 187–206.

Buus, S., Greenbaum, H., and Scharf, B. (1982). “Measurements of equal loudness and reaction times,” *J. Acoust. Soc. Am.* **72**, S94.

Buus, S., Müsch, H., and Florentine, M. (1998). “On loudness at threshold,” *J. Acoust. Soc. Am.* **104**, 399–410.

Canlon, B. (1996). “The effects of sound conditioning on the cochlea,” in *Auditory System Plasticity and Regeneration*, edited by R. Salvi, D. Henderson, F. Fiorino, and V. Colletti (Thieme, New York), pp. 128–142.

Chocholle, R. (1940). “Variation des temps de reaction auditif en fonction de l’intensité a diverses fréquences” (“Variation of auditory response times as a function of intensity at various frequencies”), *Ann. Psychol.* **41**, 65–124.

Cranford, T. W., and Krysl, P. (2015). “Fin whale sound reception mechanisms: Skull vibration enables low-frequency hearing,” *PLoS One* **10**, e0116222.

Darling, R. M., and Price, L. L. (1990). “Loudness and auditory brain stem evoked response,” *Ear Hear.* **11**, 289–295.

Davidson, S. A., Wall, L. G., and Goodman, C. M. (1990). “Preliminary studies on the use of an ABR amplitude projection procedure for hearing aid selection,” *Ear Hear.* **11**, 332–339.

de Boer, E., and de Jongh, H. R. (1978). “On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,” *J. Acoust. Soc. Am.* **63**, 115–135.

Department of Commerce and Department of the Navy (2001). “Joint interim report Bahamas marine mammal stranding event of 15–16 March 2000” (Department of Commerce, Washington, DC), p. 66.

Department of the Navy (1998). “Final environmental impact statement, shock testing the SEAWOLF submarine” (Department of the Navy, Washington, DC), p. 563.

- Department of the Navy (2001a). "Final environmental impact statement, shock trial of the WINSTON S. CHURCHILL (DDG81)" (Department of the Navy, Washington, DC), p. 598.
- Department of the Navy (2001b). "Final overseas environmental impact statement and environmental impact statement for surveillance towed array sensor system low frequency active (SURTASS LFA) sonar" (Department of the Navy, Washington, DC), p. 688.
- Department of the Navy (2008). "Appendices for the final environmental impact statement (EIS)/overseas environmental impact statement (OEIS) for the shock trial of the MESA VERDE (LPD 19)" (Department of the Navy, Washington, DC), p. 752.
- Department of the Navy (2013). "Atlantic fleet training and testing: Final environmental impact statement/overseas environmental impact statement (FEIS/OEIS)" (Department of the Navy, Washington, DC), p. 590.
- DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., Thomas, L., and Tyack, P. L. (2013). "First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar," *Biol. Letters* **9**, 20130223.
- Dolphin, W. F. (2000). "Electrophysiological measures of auditory processing in odontocetes," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 294–329.
- Dooling, R. J., Zoloth, S. R., and Baylis, J. R. (1978). "Auditory sensitivity, equal loudness, temporal resolving power, and vocalizations in the house finch (*Carpodacus mexicanus*)," *J. Comp. Physiol. Psychol.* **92**, 867–876.
- Emmerich, D. S., Gray, J. L., Watson, C. S., and Tanis, D. C. (1972). "Response latency, confidence, and ROCs in auditory signal detection," *Percept. Psychophys.* **11**, 65–72.
- Erbe, C., Reichmuth, C., Cunningham, K., Lucke, K., and Dooling, R. (2016). "Communication masking in marine mammals: A review and research strategy," *Mar. Pollut. Bull.* **103**, 15–38.
- Fay, R. R. (1988). *Hearing in Vertebrates: A Psychophysics Handbook* (Hill-Fay Associates, Winnetka, IL), p. 621.
- Fechner, G. T. (1860). *Elemente der Psychophysik (Elements of Psychophysics)* (Breitkopf and Hartel, Leipzig, Germany), p. 362.
- Finneran, J. J. (2015). "Noise-induced hearing loss in marine mammals: A review of temporary threshold shift studies from 1996 to 2015," *J. Acoust. Soc. Am.* **138**, 1702–1726.
- Finneran, J. J. (2016). "Auditory weighting functions and TTS/PTS exposure functions for marine mammals exposed to underwater noise," SSC Pacific TR 3026, San Diego, CA, p. 58.
- Finneran, J. J., Carder, D. A., Dear, R., Belting, T., McBain, J., Dalton, L., and Ridgway, S. H. (2005a). "Pure tone audiograms and possible aminoglycoside-induced hearing loss in belugas (*Delphinapterus leucas*)," *J. Acoust. Soc. Am.* **117**, 3936–3943.
- Finneran, J. J., Carder, D. A., Schlundt, C. E., and Dear, R. L. (2010a). "Growth and recovery of temporary threshold shift (TTS) at 3 kHz in bottlenose dolphins (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **127**, 3256–3266.
- Finneran, J. J., Carder, D. A., Schlundt, C. E., and Dear, R. L. (2010b). "Temporary threshold shift in a bottlenose dolphin (*Tursiops truncatus*) exposed to intermittent tones," *J. Acoust. Soc. Am.* **127**, 3267–3272.
- Finneran, J. J., Carder, D. A., Schlundt, C. E., and Ridgway, S. H. (2005b). "Temporary threshold shift (TTS) in bottlenose dolphins (*Tursiops truncatus*) exposed to mid-frequency tones," *J. Acoust. Soc. Am.* **118**, 2696–2705.
- Finneran, J. J., Dear, R., Carder, D. A., and Ridgway, S. H. (2003). "Auditory and behavioral responses of California sea lions (*Zalophus californianus*) to single underwater impulses from an arc-gap transducer," *J. Acoust. Soc. Am.* **114**, 1667–1677.
- Finneran, J. J., and Houser, D. S. (2006). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **119**, 3181–3192.
- Finneran, J. J., and Jenkins, A. K. (2012). "Criteria and thresholds for U.S. Navy acoustic and explosive effects analysis," SSC Pacific, San Diego, CA, p. 60.
- Finneran, J. J., and Schlundt, C. E. (2010). "Frequency-dependent and longitudinal changes in noise-induced hearing loss in a bottlenose dolphin (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **128**, 567–570.
- Finneran, J. J., and Schlundt, C. E. (2011). "Subjective loudness level measurements and equal loudness contours in a bottlenose dolphin (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **130**, 3124–3136.
- Finneran, J. J., and Schlundt, C. E. (2013). "Effects of fatiguing tone frequency on temporary threshold shift in bottlenose dolphins (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **133**, 1819–1826.
- Finneran, J. J., Schlundt, C. E., Branstetter, B., and Dear, R. L. (2007). "Assessing temporary threshold shift in a bottlenose dolphin (*Tursiops truncatus*) using multiple simultaneous auditory evoked potentials," *J. Acoust. Soc. Am.* **122**, 1249–1264.
- Finneran, J. J., Schlundt, C. E., Branstetter, B. K., Trickey, J., Bowman, V., and Jenkins, K. (2015). "Effects of multiple impulses from a seismic air gun on bottlenose dolphin hearing and behavior," *J. Acoust. Soc. Am.* **137**, 1634–1646.
- Finneran, J. J., Schlundt, C. E., Carder, D. A., Clark, J. A., Young, J. A., Gaspin, J. B., and Ridgway, S. H. (2000). "Auditory and behavioral responses of bottlenose dolphins (*Tursiops truncatus*) and a beluga whale (*Delphinapterus leucas*) to impulsive sounds resembling distant signatures of underwater explosions," *J. Acoust. Soc. Am.* **108**, 417–431.
- Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (2002). "Temporary shift in masked hearing thresholds (MTTS) in odontocetes after exposure to single underwater impulses from a seismic water-gun," *J. Acoust. Soc. Am.* **111**, 2929–2940.
- Fletcher, H., and Munson, W. A. (1933). "Loudness, its definition, measurement and calculation," *J. Acoust. Soc. Am.* **5**, 82–108.
- Florentine, M., Popper, A. N., and Fay, R. R. (2011). *Loudness* (Springer, New York), p. 290.
- Frantzi, A. (1998). "Does acoustic testing strand whales?," *Nature* **392**, 29.
- Gerstein, E. R., Gerstein, L., Forsythe, S. E., and Blue, J. E. (1999). "The underwater audiogram of the West Indian manatee (*Trichechus manatus*)," *J. Acoust. Soc. Am.* **105**, 3575–3583.
- Gertner, J. (2012). *The Idea Factory: Bell Labs and the Great Age of American Innovation* (Penguin Press, New York), p. 422.
- Ghoul, A., and Reichmuth, C. (2014). "Hearing in the sea otter (*Enhydra lutris*): Auditory profiles for an amphibious marine carnivore," *J. Comp. Physiol. A* **200**, 967–981.
- Glasberg, B. R., and Moore, B. C. J. (2006). "Prediction of absolute thresholds and equal-loudness contours using a modified loudness model," *J. Acoust. Soc. Am.* **120**, 585–588.
- Goldbogen, J. A., Southall, B. L., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Hazen, E. L., Falcone, E. A., Schorr, G. S., Douglas, A., Moretti, D. J., Kyburg, C., McKenna, M. F., and Tyack, P. L. (2013). "Blue whales respond to simulated mid-frequency military sonar," *Proc. R. Soc. London B* **280**, 20130657.
- Green, D. M., and Swets, J. A. (1988). *Signal Detection Theory and Psychophysics* (Peninsula Publishing, Los Altos Hills, CA), p. 487.
- Green, S. (1975). "Auditory sensitivity and equal loudness in the squirrel monkey (*Saimiri sciureus*)," *J. Exp. Anal. Behav.* **23**, 255–264.
- Hamernik, R. P., Henderson, D., Crossley, J. J., and Salvi, R. J. (1974). "Interaction of continuous and impulse noise: Audiometric and histological effects," *J. Acoust. Soc. Am.* **55**, 117–121.
- Harris, C. (1998). "Definitions, abbreviations, and symbols," in *Handbook of Acoustical Measurements and Noise Control*, edited by C. Harris (Acoustical Society of America, Woodbury, NY), pp. 2.1–2.23.
- Heffner, R. S., and Heffner, H. E. (1992). "Evolution of sound localization in mammals," in *The Evolutionary Biology of Hearing*, edited by D. B. Webster, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 691–715.
- Henderson, D., and Hamernik, R. P. (1995). "Biologic bases of noise-induced hearing loss," *Occupat. Med. State Art Rev.* **10**, 513–534.
- Henderson, D., Subramaniam, M., and Boettcher, F. A. (1993). "Individual susceptibility to noise-induced hearing loss: An old topic revisited," *Eur. Hear.* **14**, 152–168.
- Herbert, T. E., and Procter, W. S. (1934). *Telephony* (Sir Isaac Pitman and Sons, London, UK), p. 1208.
- Houser, D. S., and Finneran, J. J. (2006). "A comparison of underwater hearing sensitivity in bottlenose dolphins (*Tursiops truncatus*) determined by electrophysiological and behavioral methods," *J. Acoust. Soc. Am.* **120**, 1713–1722.
- Houser, D. S., Helweg, D. A., and Moore, P. W. B. (2001). "A bandpass filter-bank model of auditory sensitivity in the humpback whale," *Aquat. Mammal.* **27**, 82–91.
- Houser, D. S., Martin, S. W., and Finneran, J. J. (2013a). "Behavioral responses of California sea lions to mid-frequency (3250–3450 Hz) sonar signals," *Mar. Environ. Res.* **92**, 268–278.
- Houser, D. S., Martin, S. W., and Finneran, J. J. (2013b). "Exposure amplitude and repetition affect bottlenose dolphin behavioral responses to simulated mid-frequency sonar signals," *J. Exp. Mar. Biol. Ecol.* **443**, 123–133.

- Humes, L. E. (1984). "Noise-induced hearing loss as influenced by other agents and by some physical characteristics of the individual," *J. Acoust. Soc. Am.* **76**, 1318–1329.
- Jesteadt, W., Wier, C. C., and Green, D. M. (1977). "Intensity discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 169–177.
- Johnson, C. S. (1967). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon Press, Oxford), pp. 247–260.
- Johnson, C. S., McManus, M. W., and Skaar, D. (1989). "Masked tonal hearing thresholds in the beluga whale," *J. Acoust. Soc. Am.* **85**, 2651–2654.
- Kastak, D., Holt, M. M., Kastak, C. J. R., Southall, B. L., Mulsow, J., and Schusterman, R. J. (2005a). "A voluntary mechanism of protection from airborne noise in a harbor seal," in *16th Biennial Conference on the Biology of Marine Mammals*, San Diego, CA, December 12–16.
- Kastak, D., Reichmuth, C., Holt, M. M., Mulsow, J., Southall, B. L., and Schusterman, R. J. (2007). "Onset, growth, and recovery of in-air temporary threshold shift in a California sea lion (*Zalophus californianus*)," *J. Acoust. Soc. Am.* **122**, 2916–2924.
- Kastak, D., Schusterman, R. J., Southall, B. L., and Reichmuth, C. J. (1999). "Underwater temporary threshold shift induced by octave-band noise in three species of pinniped," *J. Acoust. Soc. Am.* **106**, 1142–1148.
- Kastak, D., Southall, B., Holt, M., Kastak, C. R., and Schusterman, R. (2004). "Noise-induced temporary threshold shifts in pinnipeds: Effects of noise energy," *J. Acoust. Soc. Am.* **116**, 2531–2532(A).
- Kastak, D., Southall, B. L., Schusterman, R. J., and Kastak, C. R. (2005b). "Underwater temporary threshold shift in pinnipeds: Effects of noise level and duration," *J. Acoust. Soc. Am.* **118**, 3154–3163.
- Kastelein, R. A., Bunskoek, P., Hagedoorn, M., Au, W. W. L., and de Haan, D. (2002). "Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals," *J. Acoust. Soc. Am.* **112**, 334–344.
- Kastelein, R. A., Gransier, R., and Hoek, L. (2013a). "Comparative temporary threshold shifts in a harbor porpoise and harbor seal, and severe shift in a seal," *J. Acoust. Soc. Am.* **134**, 13–16.
- Kastelein, R. A., Gransier, R., Hoek, L., Macleod, A., and Terhune, J. M. (2012a). "Hearing threshold shifts and recovery in harbor seals (*Phoca vitulina*) after octave-band noise exposure at 4 kHz," *J. Acoust. Soc. Am.* **132**, 2745–2761.
- Kastelein, R. A., Gransier, R., Hoek, L., and Olthuis, J. (2012b). "Temporary threshold shifts and recovery in a harbor porpoise (*Phocoena phocoena*) after octave-band noise at 4 kHz," *J. Acoust. Soc. Am.* **132**, 3525–3537.
- Kastelein, R. A., Gransier, R., Hoek, L., and Rambags, M. (2013b). "Hearing frequency thresholds of a harbor porpoise (*Phocoena phocoena*) temporarily affected by a continuous 1.5 kHz tone," *J. Acoust. Soc. Am.* **134**, 2286–2292.
- Kastelein, R. A., Gransier, R., Marijt, M. A. T., and Hoek, L. (2015a). "Hearing frequency thresholds of harbor porpoises (*Phocoena phocoena*) temporarily affected by played back offshore pile driving sounds," *J. Acoust. Soc. Am.* **137**, 556–564.
- Kastelein, R. A., Gransier, R., Schop, J., and Hoek, L. (2015b). "Effects of exposure to intermittent and continuous 6–7 kHz sonar sweeps on harbor porpoise (*Phocoena phocoena*) hearing," *J. Acoust. Soc. Am.* **137**, 1623–1633.
- Kastelein, R. A., Hoek, L., de Jong, C. A. F., and Wensveen, P. J. (2010). "The effect of signal duration on the underwater detection thresholds of a harbor porpoise (*Phocoena phocoena*) for single frequency-modulated tonal signals between 0.25 and 160 kHz," *J. Acoust. Soc. Am.* **128**, 3211–3222.
- Kastelein, R. A., Hoek, L., Gransier, R., Rambags, M., and Claeys, N. (2014a). "Effect of level, duration, and inter-pulse interval of 1–2 kHz sonar signal exposures on harbor porpoise hearing," *J. Acoust. Soc. Am.* **136**, 412–422.
- Kastelein, R. A., Schop, J., Gransier, R., and Hoek, L. (2014b). "Frequency of greatest temporary hearing threshold shift in harbor porpoises (*Phocoena phocoena*) depends on the noise level," *J. Acoust. Soc. Am.* **136**, 1410–1418.
- Kastelein, R. A., Schop, J., Hoek, L., and Covi, J. (2015c). "Hearing thresholds of a harbor porpoise (*Phocoena phocoena*) for narrow-band sweeps," *J. Acoust. Soc. Am.* **138**, 2508–2512.
- Kastelein, R. A., Wensveen, P., Hoek, L., and Terhune, J. M. (2009). "Underwater hearing sensitivity of harbor seals (*Phoca vitulina*) for narrow noise bands between 0.2 and 80 kHz," *J. Acoust. Soc. Am.* **126**, 476–483.
- Kastelein, R. A., Wensveen, P. J., Terhune, J. M., and de Jong, C. A. F. (2011). "Near-threshold equal-loudness contours for harbor seals (*Phoca vitulina*) derived from reaction times during underwater audiometry: A preliminary study," *J. Acoust. Soc. Am.* **129**, 488–495.
- Ketten, D. R. (2014). "Expert evidence: Chatham Rock Phosphate Ltd Application for Marine Consent," Environmental Protection Agency, New Zealand, [http://www.epa.govt.nz/EEZ/EEZ000006/EEZ000006\\_13\\_04\\_PowerPoint\\_Ketten.pdf](http://www.epa.govt.nz/EEZ/EEZ000006/EEZ000006_13_04_PowerPoint_Ketten.pdf) (Last viewed 2/16/2017).
- Ketten, D. R., and Mountain, D. C. (2014). "Inner ear frequency maps: First stage audiograms of low to infrasonic hearing in mysticetes," in *5th International Conference on the Effects of Sound in the Ocean on Marine Mammals*, Amsterdam, The Netherlands, September 7–12.
- Kohfeld, D. L., Santee, J. L., and Wallace, N. D. (1981). "Loudness and reaction time: I," *Percept. Psychophys.* **29**, 535–549.
- Kryter, K. D. (1959). "Scaling human reaction to sound from aircraft," *J. Acoust. Soc. Am.* **31**, 1415–1422.
- Kujawa, S. G., and Liberman, M. C. (2009). "Adding insult to injury: Cochlear nerve degeneration after 'temporary' noise-induced hearing loss," *J. Neurosci.* **29**, 14077–14085.
- Lei, S., Ahroon, W., and Hamerik, R. (1996). "Application of frequency and time domain kurtosis to assessment of complex, time-varying noise exposures," in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Alexson, H. Borchgrevink, R. Harmenik, H. P., D. Henderson, and R. Salvi (Thieme, New York), pp. 213–228.
- Lemons, D. W. (1999). "Auditory filter shapes in an Atlantic bottlenose dolphin (*Tursiops truncatus*)," Ph.D. thesis, Psychology Department, University of Hawaii, p. 74.
- Ljungblad, D. K., Scroggins, P. D., and Gilmartin, W. G. (1982). "Auditory thresholds of a captive Eastern Pacific bottle-nosed dolphin, *Tursiops* spp.," *J. Acoust. Soc. Am.* **72**, 1726–1729.
- Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization* (Oxford University Press, New York), p. 562.
- Lucke, K., Siebert, U., Lepper, P. A., and Blanchet, M.-A. (2009). "Temporary shift in masked hearing thresholds in a harbor porpoise (*Phocoena phocoena*) after exposure to seismic airgun stimuli," *J. Acoust. Soc. Am.* **125**, 4060–4070.
- Lutfi, R. A., and Patterson, R. D. (1984). "On the growth of masking asymmetry with stimulus intensity," *J. Acoust. Soc. Am.* **76**, 739–745.
- Mann, D., Bauer, G., Reep, R., Gaspard, J., Dziuk, K., and Read, L. (2009). *Auditory and Tactile Detection by the West Indian Manatee* (Fish and Wildlife Research Institute, St. Petersburg, FL), p. 77.
- Manoussaki, D., Chadwick, R. S., Ketten, D. R., Arruda, J., Dimitriadis, E. K., and O'Malley, J. T. (2008). "The influence of cochlear shape on low-frequency hearing," *Proc. Natl. Acad. Sci.* **105**, 6162–6166.
- Marshall, L., and Brandt, J. F. (1980). "The relationship between loudness and reaction time in normal hearing listeners," *Acta Otolaryngol.* **90**, 244–249.
- May, B. J., Little, N., and Saylor, S. (2009). "Loudness perception in the domestic cat: Reaction time estimates of equal loudness contours and recruitment effects," *J. Assoc. Res. Otolaryngol.* **10**, 295–308.
- McMinn, T. (2013). "A-weighting: Is it the metric you think it is?," in *Acoustics 2013* (Victor Harbor, Australia), pp. 1–4.
- Melnick, W., and Maves, M. (1974). "Asymptotic threshold shift (ATS) in man from 24 hour exposure to continuous noise," *Ann. Otol. Rhinol. Laryngol.* **83**, 820–828.
- Ménard, M., Gallégo, S., Berger-Vachon, C., Collet, L., and Thai-Van, H. (2008). "Relationship between loudness growth function and auditory steady-state response in normal-hearing subjects," *Hear. Res.* **235**, 105–113.
- Miller, J. (1978a). "General physiological effects of noise," in *Handbook of Perception*, edited by E. Carterette and M. Friedman (Academic, New York), pp. 677–685.
- Miller, J. (1978b). "General psychological and sociological effects of noise," in *Handbook of Perception*, edited by E. Carterette and M. Friedman (Academic, New York), pp. 641–675.
- Miller, J. D. (1978c). "Effects of noise on people," in *Handbook of Perception*, edited by E. Carterette and M. Friedman (Academic, New York), pp. 609–640.
- Miller, P. J. O., Antunes, R. N., Wensveen, P. J., Samarra, F. I. P., Catarina Alves, A., Tyack, P. L., Kvadsheim, P. H., Kleivane, L., Lam, F.-P. A., Ainslie, M. A., and Thomas, L. (2014). "Dose-response relationships for

- the onset of avoidance of sonar by free-ranging killer whales," *J. Acoust. Soc. Am.* **135**, 975–993.
- Mills, J. (1976). "Individual differences in noise-induced hearing loss," in *Hearing and Davis: Essays Honoring Hallowell Davis*, edited by S. Hirsh, D. Eldredge, I. Hirsh, and S. Silverman (Washington University Press, St. Louis, MO), pp. 115–124.
- Möhl, B. (1964). "Preliminary studies on hearing in seals," *Vidensk. Medd. Dansk. Naturh. Foren.* **127**, 283–294.
- Moody, D. B. (1970). "Reaction time as an index of sensory function," in *Animal Psychophysics: The Design and Conduct of Sensory Experiments*, edited by W. C. Stebbins (Plenum Press, New York), pp. 277–302.
- Mooney, T. A., Nachtigall, P. E., Breese, M., Vlachos, S., and Au, W. W. L. (2009a). "Predicting temporary threshold shifts in a bottlenose dolphin (*Tursiops truncatus*): The effects of noise level and duration," *J. Acoust. Soc. Am.* **125**, 1816–1826.
- Mooney, T. A., Nachtigall, P. E., and Vlachos, S. (2009b). "Sonar-induced temporary hearing loss in dolphins," *Biol. Lett.* **5**, 565–567.
- Moore, B. C. J. (1989). "The notched noise method," in *An Introduction to the Psychology of Hearing* (Academic Press, Cambridge, UK), pp. 98–102.
- Moore, B. C. J., and Glasberg, B. R. (1987). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," *Hear. Res.* **28**, 209–225.
- Mulsow, J., Houser, D. S., and Finneran, J. J. (2012). "Underwater psychophysical audiogram of a young male California sea lion (*Zalophus californianus*)," *J. Acoust. Soc. Am.* **131**, 4182–4187.
- Mulsow, J., Schlundt, C. E., Brandt, L., and Finneran, J. J. (2015). "Equal latency contours for bottlenose dolphins (*Tursiops truncatus*) and California sea lions (*Zalophus californianus*)," *J. Acoust. Soc. Am.* **138**, 2678–2691.
- Mulsow, J. L., Finneran, J. J., and Houser, D. S. (2011). "California sea lion (*Zalophus californianus*) aerial hearing sensitivity measured using auditory steady-state response and psychophysical methods," *J. Acoust. Soc. Am.* **129**(4), 2298–2306.
- Mulsow, J. L., and Reichmuth, C. (2010). "Psychophysical and electrophysiological aerial audiograms of a Steller sea lion (*Eumetopias jubatus*)," *J. Acoust. Soc. Am.* **127**, 2692–2701.
- Munk, W., and Forbes, A. M. G. (1989). "Global ocean warming: An acoustic measure?," *J. Phys. Oceanogr.* **19**, 1765–1778.
- Munk, W. H. (1991). "The Heard Island experiment," *Naval Res. Rev.* **42**, 2–22.
- Nachtigall, P. E., Lemonds, D. W., and Roitblat, H. L. (2000). "Psychoacoustic studies of dolphin and whale hearing," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 330–363.
- Nachtigall, P. E., Pawloski, J., and Au, W. W. L. (2003). "Temporary threshold shifts and recovery following noise exposure in the Atlantic bottlenose dolphin (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **113**, 3425–3429.
- Nachtigall, P. E., and Supin, A. Y. (2013). "A false killer whale reduces its hearing sensitivity when a loud sound is preceded by a warning," *J. Exp. Biol.* **216**, 3062–3070.
- Nachtigall, P. E., and Supin, A. Y. (2014). "Conditioned hearing sensitivity reduction in a bottlenose dolphin (*Tursiops truncatus*)," *J. Exp. Biol.* **217**, 2806–2813.
- Nachtigall, P. E., and Supin, A. Y. (2015). "Conditioned frequency-dependent hearing sensitivity reduction in the bottlenose dolphin (*Tursiops truncatus*)," *J. Exp. Biol.* **218**, 999–1005.
- Nachtigall, P. E., Supin, A. Y., Estaban, J. A., and Pacini, A. F. (2016). "Learning and extinction of conditioned hearing sensation change in the beluga whale (*Delphinapterus leucas*)," *J. Comp. Physiol. A* **202**, 105–113.
- Nachtigall, P. E., Supin, A. Y., Pawloski, J., and Au, W. W. L. (2004). "Temporary threshold shifts after noise exposure in the bottlenose dolphin (*Tursiops truncatus*) measured using evoked auditory potentials," *Mar. Mammal Sci.* **20**, 673–687.
- National Institute for Occupational Safety and Health (1998). *Criteria for a Recommended Standard: Occupational Noise Exposure, Revised Criteria 1998* (NIOSH, Cincinnati, OH), p. 106.
- National Marine Fisheries Service (1994). "Small takes of marine mammals incidental to shock testing the U.S.S. John Paul Jones," report No. 59 FR.
- National Marine Fisheries Service (2016). "Technical guidance for assessing the effects of anthropogenic sound on marine mammal hearing—Underwater acoustic thresholds for onset of permanent and temporary threshold shifts," report No. 81 FR 51693.
- National Oceanic and Atmospheric Administration (2013). "Draft guidance for assessing the effects of anthropogenic sound on marine mammal hearing: Acoustic threshold levels for onset of permanent and temporary threshold shifts, December 2013," NOAA, Silver Spring, MD, p. 76.
- National Oceanic and Atmospheric Administration (2015). "Draft guidance for assessing the effects of anthropogenic sound on marine mammal hearing: Underwater acoustic threshold levels for onset of permanent and temporary threshold shifts, July 2015," NOAA, Silver Spring, MD, p. 180.
- National Research Council (1994). *Low-Frequency Sound and Marine Mammals: Current Knowledge and Research Needs* (National Academy, Washington, DC), p. 75.
- Nedwell, J. R., and Turnpenny, A. W. H. (1998). "The use of a generic frequency weighting scale in estimating environmental effect," in *Proceedings of the Workshop on Seismics and Marine Mammals*, London, UK, 23–25 June.
- Nedwell, J. R., Turnpenny, A. W. H., Lovell, J., Parvin, S. J., Workman, R., Spinks, J. A. L., and Howell, D. (2007). *A Validation of the dBht as a Measure of the Behavioural and Auditory Effects of Underwater Noise* (Subacoustech Acoustic Research Consultancy, Hampshire, UK), pp. 1–74.
- Newman, J. S., and Beattie, K. R. (1985). *Aviation Noise Effects* (U.S. Department of Transportation, Federal Aviation Administration, Washington, DC), p. 105.
- Occupational Safety and Health Administration (2011). 29 CFR 1926.52, *Occupational Noise Exposure* (OSHA, Washington, DC).
- Parks, S. E., Ketten, D. R., O'Malley, J. T., and Arruda, J. (2007). "Anatomical predictions of hearing in the North Atlantic right whale," *Anat. Rec.* **290**, 734–744.
- Parmanen, J. (2012). "Some reasons to revise the International Standard ISO 226:2003: Acoustics—Normal Equal-Loudness-Level Contours," *Open J. Acoustic.* **2**, 143–149.
- Patterson, R., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1987). "An efficient auditory filterbank based on the gammatone function," in *Paper presented at the Meeting of the IOC Speech Group on Auditory Modelling at RSRE*, December 14–15.
- Patterson, R. D., and Moore, B. C. J. (1986). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 123–127.
- Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.* **72**, 1788–1803.
- Pearsons, K. S., Barber, D. S., Tabachnick, B. G., and Fidell, S. (1995). "Predicting noise-induced sleep disturbance," *J. Acoust. Soc. Am.* **97**, 331–338.
- Pfingst, B. E., Hienz, R., Kimm, J., and Miller, J. (1975a). "Reaction-time procedure for measurement of hearing. I. Suprathreshold functions," *J. Acoust. Soc. Am.* **57**, 421–430.
- Pfingst, B. E., Hienz, R., and Miller, J. (1975b). "Reaction-time procedure for measurement of hearing. II. Threshold functions," *J. Acoust. Soc. Am.* **57**, 431–436.
- Popov, V. V., Nechaev, D. I., Sysueva, E. V., Rozhnov, V. V., and Supin, A. Y. (2015). "Spectrum pattern resolution after noise exposure in a beluga whale, *Delphinapterus leucas*: Evoked potential study," *J. Acoust. Soc. Am.* **138**, 377–388.
- Popov, V. V., Supin, A. Y., Rozhnov, V. V., Nechaev, D. I., and Sysueva, E. V. (2014). "The limits of applicability of the sound exposure level (SEL) metric to temporal threshold shifts (TTS) in beluga whales, *Delphinapterus leucas*," *J. Exp. Biol.* **217**, 1804–1810.
- Popov, V. V., Supin, A. Y., Rozhnov, V. V., Nechaev, D. I., Sysueva, E. V., Klisshin, V. O., Pletenko, M. G., and Tarakanov, M. B. (2013). "Hearing threshold shifts and recovery after noise exposure in beluga whales *Delphinapterus leucas*," *J. Exp. Biol.* **216**, 1587–1596.
- Popov, V. V., Supin, A. Y., Wang, D., Wang, K., Dong, L., and Wang, S. (2011). "Noise-induced temporary threshold shift and recovery in Yangtze finless porpoises *Neophocaena phocaenoides asiaorientalis*," *J. Acoust. Soc. Am.* **130**, 574–584.
- Pratt, H., and Sohmer, H. (1977). "Correlations between psychophysical magnitude estimates and simultaneously obtained auditory nerve, brain stem and cortical responses to click stimuli in man," *Electroencephalogr. Clin. Neurophysiol.* **43**, 802–812.
- Rankovic, C. M., and Allen, J. B. (2001). *Study of Speech and Hearing at Bell Telephone Laboratories: Correspondence Files (1917–1933) and other Internal Reports* (Acoustical Society of America, Melville, NY), CD-ROM.

- Reed, C. M., and Bilger, R. C. (1973). "A comparative study of S/N0 and E/N0," *J. Acoust. Soc. Am.* **53**, 1039–1044.
- Reichmuth, C. (2013). "Equal loudness contours and possible weighting functions for pinnipeds," *J. Acoust. Soc. Am.* **134**, 4210(A).
- Reichmuth, C., Holt, M. M., Mulsow, J., Sills, J. M., and Southall, B. L. (2013). "Comparative assessment of amphibious hearing in pinnipeds," *J. Comp. Physiol. A* **199**, 491–507.
- Reichmuth, C., and Southall, B. L. (2012). "Underwater hearing in California sea lions (*Zalophus californianus*): Expansion and interpretation of existing data," *Mar. Mammal Sci.* **28**, 358–363.
- Ridgway, S., and Carder, D. (2000). "A preliminary study of loudness at frequencies of 5 to 120 kHz based on whistle response time (RT) in a dolphin," *J. Acoust. Soc. Am.* **108**, 2515(A).
- Ridgway, S. H., Carder, D. A., Kamolnick, T., Smith, R. R., Schlundt, C. E., and Elsberry, W. R. (2001). "Hearing and whistling in the deep sea: Depth influences whistle spectra but does not attenuate hearing by white whales (*Delphinapterus leucas*) (Odontoceti, Cetacea)," *J. Exp. Biol.* **204**, 3829–3841.
- Ridgway, S. H., Carder, D. A., Smith, R. R., Kamolnick, T., Schlundt, C. E., and Elsberry, W. R. (1997). *Behavioral Responses and Temporary Shift in Masked Hearing Thresholds of Bottlenose Dolphins, Tursiops truncatus, to 1-second Tones of 141–201 dB re 1  $\mu$ Pa* (Naval Command, Control, and Ocean Surveillance Center, RDT&E Division, San Diego, CA), p. 31.
- Riesz, R. R. (1928). "Differential intensity sensitivity of the ear for pure tones," *Phys. Rev.* **31**, 867–875.
- Robinson, D. W., and Dadsen, R. S. (1956). "A re-determination of the equal-loudness relations for pure tones," *Br. J. Appl. Phys.* **7**, 166–181.
- Royster, J. D. (1996). "Noise-induced hearing loss," in *Hearing Disorders*, edited by J. Northern (Allyn and Bacon, Boston, MA), pp. 177–188.
- Schlundt, C. E., Dear, R. L., Green, L., Houser, D. S., and Finneran, J. J. (2007). "Simultaneously measured behavioral and electrophysiological hearing thresholds in a bottlenose dolphin (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **122**, 615–622.
- Schlundt, C. E., Finneran, J. J., Branstetter, B. K., Dear, R. L., Houser, D. S., and Hernandez, E. (2008). "Evoked potential and behavioral hearing thresholds in nine bottlenose dolphins (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **123**, 3506(A).
- Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (2000). "Temporary shift in masked hearing thresholds of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones," *J. Acoust. Soc. Am.* **107**, 3496–3508.
- Schuknecht, H. (1974). *Pathology of the Ear* (Harvard University Press, Cambridge, MA), p. 444.
- Serpanos, Y. C., O'Malley, H., and Gravel, J. S. (1997). "The relationship between loudness intensity functions and the click-ABR wave V latency," *Ear Hear.* **18**, 409–419.
- Shower, E. G., and Biddulph, R. (1931). "Differential pitch sensitivity of the ear," *J. Acoust. Soc. Am.* **3**, 275–287.
- Sills, J. M., Southall, B. L., and Reichmuth, C. (2014). "Amphibious hearing in spotted seals (*Phoca largha*): Underwater audiograms, aerial audiograms and critical ratio measurements," *J. Exp. Biol.* **217**, 726–734.
- Sills, J. M., Southall, B. L., and Reichmuth, C. (2015). "Amphibious hearing in ringed seals (*Pusa hispida*): Underwater audiograms, aerial audiograms and critical ratio measurements," *J. Exp. Biol.* **218**, 2250–2259.
- Silva, I., and Epstein, M. (2010). "Estimating loudness growth from tone-burst evoked responses," *J. Acoust. Soc. Am.* **127**, 3629–3642.
- Sivian, L. J., and White, S. D. (1933). "On minimum audible sound fields," *J. Acoust. Soc. Am.* **4**, 288–321.
- Sivle, L. D., Kvadsheim, P. H., Fahlman, A., Lam, F. P. A., Tyack, P. L., and Miller, P. J. O. (2012). "Changes in dive behavior during naval sonar exposure in killer whales, long-finned pilot whales, and sperm whales," *Front. Physiol.* **3**, 400.
- Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene, C. R., Jr., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., Richardson, W. J., Thomas, J. A., and Tyack, P. L. (2007). "Marine mammal noise exposure criteria: Initial scientific recommendations," *Aquat. Mammal.* **33**, 411–521.
- Stebbins, W. C. (1966). "Auditory reaction time and the derivation of equal loudness contours for the monkey," *J. Exp. Anal. Behav.* **9**, 135–142.
- Steinberg, J. C., Montgomery, H. C., and Gardner, M. B. (1940a). "Results of the World's Fair hearing tests," *J. Acoust. Soc. Am.* **12**, 291–301.
- Steinberg, J. C., Montgomery, H. C., and Gardner, M. B. (1940b). "Results of World's Fair hearing tests," *Bell Syst. Tech. J.* **19**, 533–562.
- Stevens, S. S. (1946). "On the theory of scales of measurement," *Science* **103**, 677–680.
- Subramaniam, M., Henderson, D., and Henselman, L. (1996). "'Toughening' of the mammalian auditory system: Spectral, temporal, and intensity effects," in *Auditory System Plasticity and Regeneration*, edited by R. Salvi, D. Henderson, F. Fiorino, and V. Colletti (Thieme, New York), pp. 128–142.
- Suter, A. (1996). "Current standards for occupational noise exposure," in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Alexson, H. Borchgrevink, R. Harmenik, H. P. D. Henderson, and R. Salvi (Thieme, New York), pp. 430–436.
- Suzuki, Y., and Takeshima, H. (2004). "Equal-loudness-level contours for pure tones," *J. Acoust. Soc. Am.* **116**, 918–933.
- Terhune, J. M. (1988). "Detection thresholds of a harbour seal to repeated underwater high-frequency, short-duration sinusoidal pulses," *Can. J. Zool.* **66**, 1578–1582.
- Terhune, J. M. (2013). "A practical weighting function for harbor porpoise underwater sound level measurements," *J. Acoust. Soc. Am.* **134**, 2405–2408.
- Thornton, A., Farrell, G., and McSparran, E. (1989). "Clinical methods for the objective estimation of loudness discomfort level (LDL) using auditory brainstem responses in patients," *Int. J. Audiol.* **18**, 225–230.
- Thornton, A., Yardley, L., and Farrell, G. (1987). "The objective estimation of loudness discomfort level using auditory brainstem responses," *Scand. Audiol.* **16**, 219–225.
- Tougaard, J., Wright, A. J., and Madsen, P. T. (2015). "Cetacean noise criteria revisited in the light of proposed exposure limits for harbour porpoises," *Mar. Pollut. Bull.* **90**, 196–208.
- Tubelli, A. A., Zosuls, A., Ketten, D. R., Yamato, M., and Mountain, D. C. (2012). "A prediction of the minke whale (*Balaenoptera acutorostrata*) middle-ear transfer function," *J. Acoust. Soc. Am.* **132**, 3263–3272.
- Viemeister, N. F., and Plack, C. J. (1993). "Time Analysis," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 116–154.
- Voipio, H.-M., Nevalainen, T., Halonen, P., Hakumäki, M., and Björk, E. (2006). "Role of cage material, working style and hearing sensitivity in perception of animal care noise," *Lab. Anim.* **40**, 400–409.
- Wagner, E., Florentine, M., Buus, S., and McCormack, J. (2004). "Spectral loudness summation and simple reaction time," *J. Acoust. Soc. Am.* **116**, 1681–1686.
- Ward, W. D. (1976). "A comparison of the effects of continuous, intermittent, and impulse noise," in *Effects of Noise on Hearing*, edited by D. Henderson, R. P. Hamernik, D. S. Dosanjh, and J. H. Mills (Raven Press, New York), pp. 407–420.
- Wegel, R., and Lane, C. E. (1924). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," *Phys. Rev.* **23**, 266–285.
- Wensveen, P. J., Huijser, L. A. E., Hoek, L., and Kastelein, R. A. (2014). "Equal latency contours and auditory weighting functions for the harbour porpoise (*Phocoena phocoena*)," *J. Exp. Biol.* **217**, 359–369.
- White, M. J., Norris, J., Ljungblad, D. K., Baron, K., and di Sciara, G. N. (1978). *Auditory Thresholds of Two Beluga Whales (Delphinapterus leucas)* (Hubbs Sea World Research Institute, San Diego, CA), p. 36.
- Wier, C. C., Jesteadt, W., and Green, D. M. (1977). "Frequency discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 178–184.
- Wilson, K. G., and Stelmack, R. M. (1982). "Power functions of loudness magnitude estimations and auditory brainstem evoked responses," *Percept. Psychophys.* **31**, 561–565.
- Worcester, P. F., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Howe, B. M., Mercer, J. A., Spindel, R. C., Colosi, J. A., Metzger, K., Birdsall, T. G., and Baggeroer, A. B. (1999). "A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean," *J. Acoust. Soc. Am.* **105**, 3185–3201.
- Yeager, D., and Marsh, A. (1998). "Sound levels and their measurement," in *Handbook of Acoustical Measurements and Noise Control*, edited by C. Harris (Acoustical Society of America, Woodbury, NY), pp. 11.11–11.18.
- Yost, W. A. (2015). "Psychoacoustics: A Brief Historical Overview," *Acoust. Today* **11**, 46–53.
- Yost, W. A., and Killion, M. C. (1997). "Hearing thresholds," *Encycl. Acoust.* **3**, 1545–1554.
- Yost, W. A., Popper, A. N., and Fay, R. R. (1993). *Human Psychoacoustics* (Springer-Verlag, New York), p. 242.
- Yost, W. A., and Shofner, W. P. (2009). "Critical bands and critical ratios in animal psychoacoustics: An example using chinchilla data," *J. Acoust. Soc. Am.* **125**, 315–323.
- Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Y. (2005). "Behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*)," *J. Acoust. Soc. Am.* **118**, 2688–2695.