Aquatic Mammals 2007, *33*(1), 43-54, DOI 10.1578/AM.33.1.2007.43

Objective Detection of Bottlenose Dolphin (*Tursiops truncatus*) Steady-State Auditory Evoked Potentials in Response to AM/FM Tones

James J. Finneran,[1] Dorian S. Houser,[2] and Carolyn E. Schlundt[3]

[1]*U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, CA 92152, USA;*
E-mail: james.finneran@navy.mil
[2]*BIOMIMETICA, 7951 Shantung Drive, Santee, CA 92071-3432, USA*
[3]*EDO Professional Services, 3276 Rosecrans Street, San Diego, CA 92110, USA*

Abstract

Auditory steady-state responses were measured in a bottlenose dolphin (*Tursiops truncatus*) and used to illustrate objective techniques to determine the presence or absence of a response. Experimental measurements were conducted under water in a quiet pool. Sound stimuli were pure tones that were both amplitude and frequency modulated. Evoked responses were recorded using noninvasive surface electrodes. Two frequency-domain techniques were used to assess the presence or absence of a response. The *F* test compares the evoked potential power at a single frequency (the amplitude modulation frequency) to the noise power averaged over adjacent frequencies. Magnitude-squared coherence (MSC) is a ratio of the signal power at a single frequency to the signal-plus-noise power and reflects the degree to which the system output is determined by the input. For the measurements here, both techniques provided identical results. Evoked potential thresholds based on the lowest detected response compared favorably to behavioral thresholds obtained in the same environment.

Key Words: hearing, evoked potentials, objective detection, bottlenose dolphin, *Tursiops truncatus*

Introduction

Auditory evoked potentials (AEPs) have become increasingly popular for measurements of auditory capabilities, particularly in infants, animal subjects, and other populations that are difficult to test or to whom access is limited. One of the primary applications of AEP measurements has been to estimate hearing thresholds—measures of the quietest sounds that can be detected. Although AEP hearing thresholds may be estimated using a variety of sound stimuli, the use of sinusoidally amplitude-modulated (SAM) tones offers certain advantages. SAM tone stimuli produce the so-called "envelope-following response" (EFR) or "auditory steady-state response" (ASSR)—a harmonic evoked potential with a fundamental frequency at the stimulus modulation frequency (Campbell et al., 1977; Hall, 1979; Stapells et al., 1984; Kuwada et al., 1986; Dolphin & Mountain, 1992). Because SAM tones may possess relatively narrow frequency bandwidth and the ASSR may be analyzed in the frequency domain, ASSR measurements are commonly used for assessments of frequency-dependent hearing thresholds in both marine and terrestrial mammals (e.g., Rickards & Clark, 1984; Dolphin & Mountain, 1992; Dolphin, 1995; Rance et al., 1995; Supin & Popov, 1995; Lins et al., 1996; Supin et al., 2001; Nachtigall et al., 2005; Yuen et al., 2005; Cook et al., 2006).

ASSR thresholds are typically estimated by performing a series of AEP measurements at various stimulus sound pressure levels (SPLs), including SPLs low enough that the evoked response is no longer judged to be present. Thresholds are normally defined using either the lowest stimulus level producing a detectable response (e.g., Lins et al., 1996; Vander Werff & Brown, 2005) or interpolating/extrapolating from the measured data to determine the stimulus level corresponding to an arbitrary ASSR amplitude—often zero (e.g., Campbell et al., 1977; Supin et al., 2001; Nachtigall et al., 2004). The decision as to whether a response is present or absent has traditionally been made by a human observer. The observer may examine time domain waveforms or frequency spectra, but in either case, the performance can be highly variable (Dobie & Wilson, 1993). Problems with inter- and intraobserver reliability result from difficulties human observers experience when attempting to detect AEPs near threshold (Rose et al., 1971; Gans et al., 1992; Dobie & Wilson, 1995). In addition, any biases the observer may have, such as that resulting from the observer's experience or *a priori* knowledge of the subject's behavioral threshold, may affect the resulting ASSR threshold.

To eliminate problems related to reliability and observer bias, a number of objective response detection (ORD) techniques have been applied to ASSR detection. These techniques allow false positive rates to be explicitly specified and may, in some cases, have superior performance to human observers. In addition, ORD techniques may not require a trained observer and may therefore allow a greater degree of computer control and a more rapid testing pace to be achieved. Even in cases where ASSR presence or absence is still determined by human observers, ORD methods may provide useful information (Dobie & Wilson, 1993).

ORD techniques can be performed in the time domain or in the frequency domain. Time domain methods compare AEPs to some predefined response template or noise estimates based on amplitude or power ratios. Frequency domain methods compare the AEP amplitude in some specified frequency band to a noise estimate obtained either from a control trial or concurrently with the AEP measurement (Dobie & Wilson, 1993, 1995). Frequency domain ORD techniques are particularly attractive for use with the ASSR since it possesses a known fundamental frequency (Dobie & Wilson, 1994a).

Frequency-domain ORD methods include the *F* test (Dobie & Wilson, 1996; Lins et al., 1996), magnitude-squared coherence (MSC) (Dobie & Wilson, 1989), phase coherence (PC) (Jerger, 1986; Picton et al., 1987a; Stapells et al., 1987), Hotelling T^2 test (HT^2) (Hotelling, 1931), and the circular T^2 test (CT^2) (Victor & Mast, 1991). The *F* test evaluates whether the power at a particular frequency is statistically different from the noise power averaged over adjacent frequencies (Dobie & Wilson, 1996; Lins et al., 1996). MSC, PC, HT^2, and CT^2 require the AEP data to be segmented into multiple "subaverages." MSC represents a ratio of signal power to signal plus noise power (Tucci et al., 1990; Dobie & Wilson, 1989, 1993, 1995, 1996). PC, also known as the Rayleigh statistic or vector strength, measures the degree to which the phase angles in a number of subaverages are clustered (indicating a response) or randomly dispersed (no response) (Dobie & Wilson, 1989, 1993). PC is analogous to the square-root of MSC but disregards amplitude information and uses only phase information (Dobie & Wilson, 1989). HT^2 is a bivariate version of the Student's *t* test and calculates a two-dimensional confidence ellipse for the subaverage response vectors; if the ellipse does not include the origin, the response is considered present (Picton et al., 1987b; Dobie & Wilson, 1993, 1994b; Lins et al., 1995, 1996). CT^2 is a biased estimate of the signal-to-noise power ratio and has been shown to be algebraically related to MSC (Dobie & Wilson, 1993). Although CT^2 has been demonstrated to be superior to HT^2, Dobie & Wilson (1996) suggest that the *F* test, MSC, and CT^2 are equivalent in statistical power. Comparisons between MSC and PC have been equivocal; for example, MSC has performed better under conditions of a fixed signal added to noise at various signal-to-noise ratios (SNRs), but PC has performed better under conditions of nonstationary noise (Dobie & Wilson, 1994b).

Although frequency-domain ORD techniques, such as PC and MSC, are commonly used with human ASSR measurements (Picton et al., 1987a, 1987b; Stapells et al., 1987; Dobie & Wilson, 1989, 1993, 1994b, 1996; Lins et al., 1995, 1996), there are few examples of their use with marine mammals. Dolphin (2000) used a statistical technique to compare ASSR power at the modulation frequency to noise power at nearby frequencies. Cook et al. (2006) performed *ad hoc* comparisons of ASSR spectral amplitude to noise spectral amplitudes in a beaked whale (*Mesoplodon europaeus*), though this technique did not allow explicit specification of false positive rates. MSC was used in ASSR threshold measurements in bottlenose dolphins (*Tursiops truncatus*) (Finneran & Houser, 2006, 2007; Houser & Finneran, 2006a, 2006b) and northern elephant seals (*Mirounga angustirostris*) (Houser et al., this issue).

This paper discusses the application of two frequency-domain ORD techniques—the *F* test and MSC—to measurements of hearing sensitivity in a bottlenose dolphin. The ORD techniques are reviewed and the hardware measurement system is described in some detail. The resulting ASSR thresholds are compared to behavioral thresholds obtained in the same environment.

Materials and Methods

ORD Techniques

Frequency-domain ORD techniques for ASSR measurements are essentially statistical methods for detecting the presence of a sinusoid of a known frequency f_s. For ASSR measurements, f_s is typically the fundamental frequency of the evoked response (but see Campbell et al., 1977, for ASSR analysis at twice-the-modulation frequency).

The *F* test is a technique for comparing power estimates at f_s to noise power at frequencies adjacent to f_s. The power measured at f_s, P_{s+n}, is an unbiased estimate of the sum of the signal power and the noise power. The average power across "*m*" frequencies adjacent to f_s, P_n, is an unbiased estimate of the noise power (Dobie & Wilson, 1996). Since both power at f_s and neighboring frequencies are distributed as *chi*-square (Zurek, 1992), their ratio can be tested using an *F* statistic:

$$F = \frac{P_{s+n}}{P_n}. \qquad (1)$$

At each frequency, measured power is the sum of two independent variables (the real and imaginary parts of the complex amplitude), so the degrees of freedom in P_{s+n} and P_n are 2 and 2*m*, respectively. The ratio *F* may therefore be tested for statistical significance using standard tables to obtain the critical value for *F* with 2, 2*m* degrees of freedom: $F_{crit} = F_{(2, 2m)}$ (Dobie & Wilson, 1996).

Performance of the *F* test improves with increasing *m*; however, there are diminishing returns for $m > 15$ (Dobie & Wilson, 1996). Performance will also increase with increasing epoch (record) length since this reduces the size of the frequency bins and the resulting noise power within each bin (assuming noise spectral density is flat across frequencies). There are practical limits to epoch size, however, since longer epochs not only require more time for data collection and analysis but may also interfere with the ability to reject epochs containing large artifacts. Also, frequency analysis bin size cannot be reduced to the point where the bins are smaller than the finite bandwidth of the ASSR at f_s since the signal would then bias the noise distribution unless adjacent frequency bins were excluded.

MSC relies on segment analysis, where the collection of recorded epochs is divided into a number of individual segments, each of which is then averaged (either synchronously in the time domain or coherently in the frequency domain) to yield a number of "subaverages," each derived from a unique subset of the original collection of epochs. The subaverages are also combined to produce a "grand average."

MSC is the quantity normally referred to as "coherence" in signal processing applications and is a measure of the degree to which a system's output is determined by the input (Bendat & Piersol, 1986). For ASSR measurements, the coherence calculation is simplified because the system input is periodic and time-locked to the data collection period, and spectral estimates are averaged across segments (Dobie & Wilson, 1989). Under these conditions, MSC may be calculated from

$$MSC = \frac{\left| \frac{1}{Q} \sum_{q=1}^{Q} Y_q(f_s) \right|^2}{\frac{1}{Q} \sum_{q=1}^{Q} \left| Y_q(f_s) \right|^2} \qquad (2)$$

where *Q* is the number of subaverages, and $Y_q(f_s)$ is the complex spectral amplitude at f_s for the *q*th subaverage. Examination of Equation (2) reveals that MSC is a ratio of the power in the grand average ("power of the mean"—PM) to the average power of the subaverages ("mean power"—MP). The grand average is created entirely from coherent averaging and, thus, the noise tends to cancel out (Dobie & Wilson, 1996). The noise does not similarly cancel when the powers of the subaverages are averaged together (this is a form of rms averaging which does not increase signal-to-noise), so this quantity represents "signal plus noise." MSC is therefore a ratio of signal to signal-plus-noise and varies from 0 (all noise) to 1 (all signal) (Dobie & Wilson, 1989). Multiple response spectra are required for smoothing to create a valid coherence estimate from the AEP data (Dobie & Wilson, 1989), hence the need for segment analysis. Increasing the number of subaverages will improve detection; however, there is a point of diminishing returns when $Q > 16$ (Dobie & Wilson, 1996). Critical values for MSC are available from several authors (e.g., Amos & Koopmans, 1963; Brillinger, 1978).

Subject and Test Environment

A series of AEP measurements was conducted to illustrate the application of Equations (1 & 2). The subject (BLU) was a 41-y-old, 200-kg female bottlenose dolphin with extensive experience in cooperative psychophysical tasks and evoked potential recordings. The subject was housed in floating netted enclosures (9×9 to 12×24 m) located in San Diego Bay, California. Experimental measurements were conducted in an above-ground, vinyl-walled, seawater-filled pool approximately 3.7×6×1.5 m. Ambient noise levels in both environments were measured using a Reson TC-4032 low-noise hydrophone (Figure 1). The curve for San Diego Bay is the average over a 30-d period. The data in the pool were collected in a single session (100 averages). Ambient noise levels in the pool were 25 to 30 dB below those in San Diego Bay over the frequency range 0.1 to 4 kHz. At frequencies above approximately 4 kHz, the pool ambient noise levels were below the self-noise of the measurement system.

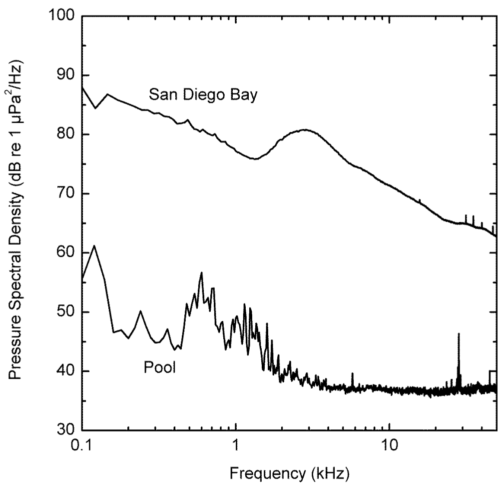

Figure 1. Ambient noise levels measured in the test pool at San Diego Bay

The subject was trained to position on a plastic "biteplate" at a depth of approximately 14 cm. The depth was chosen to keep the top of the subject's head above water. The study followed a protocol approved by the Institutional Animal Care and Use Committee at the Space and Naval Warfare Systems Center, San Diego, and followed all applicable U.S. Department of Defense guidelines.

Electrophysiological Measurements

The measurement system was centered on a rugged, portable notebook computer (Dolch NotePAC), featuring an expansion chassis with two ¾-length PCI slots (Figure 2). A multifunction data acquisition board (National Instruments PCI-6251) resided in one PCI slot and was used to generate sound stimuli with 16-bit resolution at a rate of 2 MHz. Outgoing stimuli were filtered from 0.2 to 150 kHz (Krohn-Hite 3C series) and passed through a custom programmable attenuator. The attenuator was controlled using digital outputs from the PCI-6251 and featured three stages (10, 20, and 35 dB), allowing attenuation of up to 65 dB and a usable output dynamic range of > 110 dB. The filter and attenuator were housed within a second expansion chassis attached to the computer. Stimuli were transmitted using a spherical piezoelectric sound projector (ITC 1032) driven from the attenuator output.

Sound stimuli were continuously generated during the evoked potential recordings (there were no temporal gaps during a single AEP measurement). Stimuli consisted of sinusoidal tones employing both frequency modulation (FM) and amplitude modulation (AM). Although AM tones are often used to elicit the ASSR, for this particular experiment, FM was also employed as a means

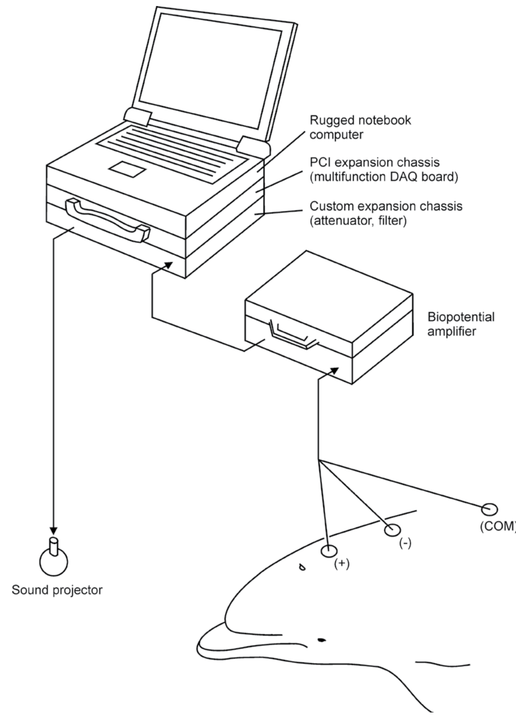

Figure 2. AEP measurement system components; the electrodes are represented by the ovals labeled (+), (-), and (COM).

of creating more uniform sound pressures within the test pool (Kastelein et al., 2002; Finneran & Schlundt, 2006). The relatively small dimensions of the pool resulted in complex pure-tone sound fields and variations in AM depth due to small changes in subject position; the use of FM stimuli creates a more uniform overall sound pressure since the effects of constructive and destructive interference tend to cancel over a number of frequencies. Combination AM/FM tones have been previously used in human ASSR measurements, primarily to increase the ASSR amplitude compared to AM only (Picton et al., 1987b; Cohen et al., 1991; Dimitrijevic et al., 2002).

Stimulus center frequencies were 10, 20, 30, 40, 50, 60, and 80 kHz. The FM was sinusoidal at a rate of 100 Hz with a 10% frequency bandwidth (relative to the center frequency [f_c]). The AM was also sinusoidal with a depth of 100% and a rate of 1 kHz. Previous sound field measurements in the pool showed that an FM bandwidth of 10% provided dramatic improvement in the measured sound fields over regions comparable in size to the subject's head (Finneran & Schlundt, 2006). Measurements of auditory filter shapes in *Tursiops* at 20, 30, and 40 kHz suggest that a bandwidth of 10% would be less than the auditory filter equivalent rectangular bandwidth (ERB) (Lemonds et al., 1997; Finneran et al., 2002), so calibration measurements based on the total SPL over the entire stimulus bandwidth would not tend to overestimate the actual received stimulus. Previous measurements in *Tursiops* (Dolphin et al., 1995; Supin & Popov, 1995) and BLU in particular (Finneran & Houser, 2006) have shown AM tones with 1 kHz modulation rate produce relatively large evoked responses.

This particular AM function produced a single peak at the f_c and sidebands located at the f_c ± the AM frequency (f_{AM} = 1 kHz). The FM added a 10% bandwidth (relative to the f_c) to each of the AM components. The resulting stimulus therefore contained a range of frequencies spanning ($0.95f_c - f_{AM}$) to ($1.05f_c + f_{AM}$). Sound stimuli were calibrated in the pool using a hydrophone (B&K 8105) positioned at the location corresponding to the midpoint between the subject's ears (without the subject present). The hydrophone output was amplified and filtered (B&K 2635) then digitized by the PCI-6251. Frequency analysis was performed on 39-ms records to obtain the frequency spectra of the various AM/FM tones (Figure 3). Comparison measurements made using the hydrophone positioned near the subject's lower jaw while on the biteplate were within 3 dB of the calibrated values.

AEPs were measured using 10-mm gold cup electrodes embedded in silicon rubber suction cups. Electrode signals were amplified (10^5 gain) and filtered (0.3 to 3 kHz) using a differential biopotential amplifier (Grass ICP-511). A three-electrode configuration was used (see Figure 2) with the noninverting (+) electrode located near the midline close to the blowhole, the inverting (-) electrode positioned along the midline

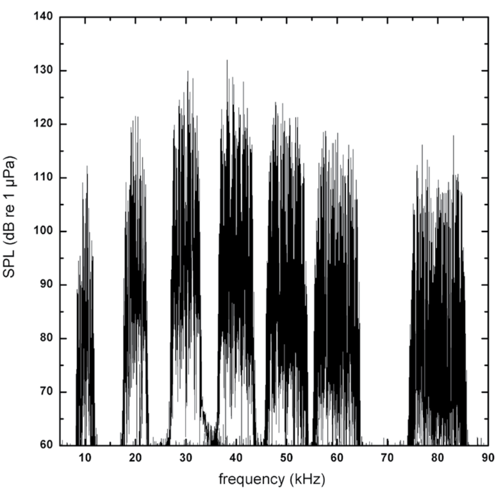

Figure 3. Frequency spectra for combination AM/FM sound stimuli

approximately 30 cm posterior to the noninverting electrode, and a common electrode located on the dorsal fin. Creases in the subject's head prevented a more optimal location for the noninverting electrode along the midline posterior to the blowhole (as described in Popov & Supin, 1990).

The biopotential amplifier output was digitized at 10 kHz using the PCI-6251. Signals were continuously acquired in 39-ms epochs. Epochs with a peak instantaneous voltage exceeding 13 µV were rejected from the analysis. The first two epochs were also always rejected to allow the sound pressure to reach a steady-state within the pool.

At a 30 kHz f_c, a single collection of ASSR measurements was made over a range of SPLs from 52 to 132 dB re 1 µPa in 5-dB steps. These data were more comprehensive than those resulting from threshold testing (see below) and were used to illustrate some features of the ORD techniques. At each SPL, 1,000 epochs were acquired. MSC was calculated at f_{AM} using 20 subaverages and an acceptable false positive rate (α) of 0.01. Each subaverage was derived from 50 sequential epochs, such that the first subaverage was calculated from epochs 1 through 50, the second from epochs 51 through 100, etc. Critical values for MSC were obtained from Amos & Koopmans (1963) and Brillinger (1978). To perform the *F* test, the 20 subaverages were sequentially appended and analyzed (in the frequency domain) as a single record with a duration of 780 ms. This resulted in a frequency resolution of 1.28 Hz. The *F* test was performed by comparing the power at f_{AM} to the average power in 19 adjacent frequencies (9 below and 10 above), a total bandwidth of 24.3 Hz about the 1,000 Hz fundamental. The critical value for *F* was calculated using standard statistical functions.

During threshold testing, the presence or absence of an evoked response was determined after 500 epochs were collected by comparing the MSC at f_{AM} to the critical value for MSC using 20 subaverages and α = 0.01. If a response was not detected, an additional 500 epochs were collected (1,000 total); if a response was detected, the measurement was complete after 500 epochs.

At each center frequency, threshold testing began at an SPL of approximately 120 dB re 1 µPa. A modified staircase technique was used to adjust the stimulus SPL after each trial. If an evoked response was detected, then the SPL for the next measurement was reduced by the step size ΔL; if a response was not detected, then the SPL was increased by ΔL. The step size was adjusted after each transition (reversal) from a detection to a non-detection, or vice versa, according to the rules:

$$\Delta L_{n+1} = 0.4\,\Delta L_n$$

(for reversals following detections),

$$\Delta L_{n+1} = 0.45\,\Delta L_n$$

(for reversals following nondetections), (3)

where ΔL_n is the step size for the nth measurement. The intent of the staircase technique was to approach threshold while avoiding, if possible, repeated testing at any particular SPLs (for this reason, multipliers of 0.4 and 0.45 were used rather than 0.5). The starting step size (ΔL_1) was 30 dB except at 80 kHz where 15 dB was used instead. The staircase was terminated when the step size for the next measurement was < 3 dB. The threshold was defined as the mean of the stimulus SPLs corresponding to the lowest hit and the next highest miss. Detections occurring at SPLs below two nondetections were considered false positives and were excluded from threshold calculations. Thresholds from three sessions were averaged to yield a mean threshold estimate at each center frequency.

Behavioral Measurements

The behavioral approach was based on the Method of Free Response, or MFR (Egan et al., 1961), and is described in detail by Finneran et al. (2005). Measurements were made using the same biteplate apparatus used in the AEP measurements. Hearing test tones were 500 ms in duration with 50 ms rise and fall times. Tones were modulated using linear FM from 0.95 f_c to 1.05 f_c (10% bandwidth). Preliminary data from the pool with BLU showed hearing thresholds obtained with linear FM and sinusoidal FM with 10% bandwidths to be typically within 5 dB (Finneran & Schlundt, 2006). An underwater light was used to delineate the trials, of which 50% were no-tone ("catch") trials. The subject was trained to whistle in response to tones and stay quiet otherwise. Stimulus SPLs were adjusted using a descending staircase technique (Cornsweet, 1962) with a 2-dB step size. Threshold estimates were based on six consecutive hit-miss or miss-hit reversals. Thresholds from three independent sessions were averaged to yield a mean threshold estimate.

Results

Time waveforms and frequency spectra for ASSRs to 30 kHz f_c AM/FM tones are shown in Figures 4 and 5. At high SPLs, evoked response waveforms exhibited a sinusoidal pattern with some amplitude modulation, presumably caused by the change in received stimulus SPL as the frequency swept through a series of maxima/minima. With an FM rate of 100 Hz, about four cycles of FM occurred within the 39-ms epoch, which is confirmed by the response waveforms. At lower SPLs, the response amplitudes diminished and the AM nature of the responses became less apparent. ASSR frequency spectra possess sidebands about the fundamental

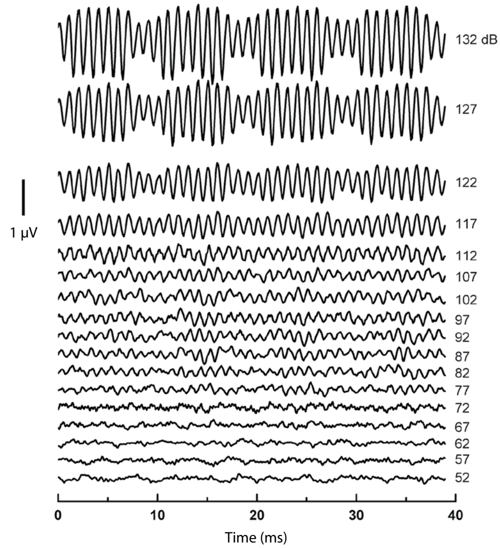

Figure 4. ASSR time waveforms in response to 30 kHz f_c AM/FM stimuli; the numbers in the right margin indicate the SPL for each trace.

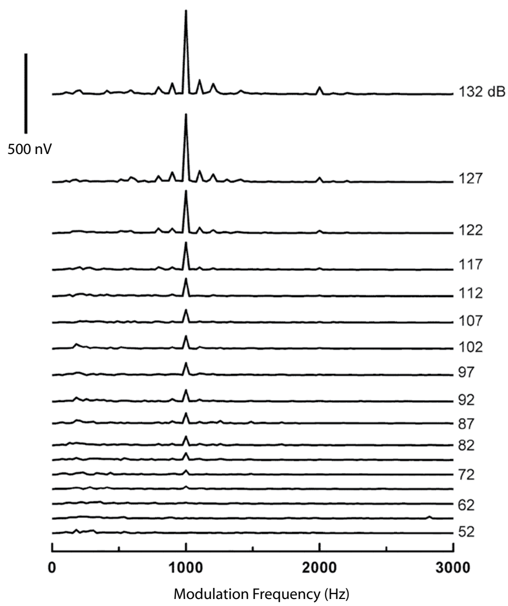

Figure 5. ASSR frequency spectra generated by 30 kHz f_c AM/FM stimuli; the numbers in the right margin indicate the SPL for each trace.

at $f_{AM} \pm f_{FM}$, $f_{AM} \pm 2f_{FM}$, confirming the amplitude modulation at high SPLs caused by the received SPLs fluctuating at the FM rate. Second harmonics at $2f_{AM}$ were also visible at high SPLs.

The amplitude and phase angle at f_{AM} for each of the spectra shown in Figure 5 are shown in Figure 6. The amplitude pattern exhibits a familiar form with an approximately linear relationship between ASSR amplitude and SPL at low stimulus levels, a plateau region where changes in SPL have little effect on ASSR amplitude, followed by a region where ASSR amplitude increases steeply with stimulus SPL. Phase angles varied linearly across most of the range of stimulus levels.

Examples of power spectra used in the *F* test calculation are shown in Figure 7 for stimulus SPLs of 52, 62, 82, and 102 dB re 1 µPa (note the differences in vertical scales). The relationship between the spectral peak at f_{AM} and the nearby noise gives a visual estimate of the relationship between signal and signal-plus-noise. For a more quantitative comparison, the spectral peak at f_{AM} must be compared to the noise levels and F_{crit}. The spectral value at 1,000 Hz ($=f_{AM}$) is P_{s+n} and the thin dashed line represents the noise power estimate, P_n (the average power at m = 19 frequencies adjacent to f_{AM}). The thick dashed line represents $F_{crit}{\cdot}P_n$; if P_{s+n} exceeds this value, then $F > F_{crit}$ and a response is considered present (102, 82, and 62 dB re 1 µPa plots), otherwise no response is detected (52 dB re 1 µPa plot). *F* ratio values calculated for the entire data set (including the data of Figure 7) are shown in Figure 8 as a function of stimulus SPL. The *F* ratio generally increased with stimulus SPL; however, there are some dips in the *F* ratio between 100 to 120 dB re 1 µPa where the ASSR amplitude plateaus. When the SPL ≥ 62 dB re 1 µPa, *F* ratios were above the value of F_{crit} (dashed line) indicating detected responses (m = 19, α = 0.01).

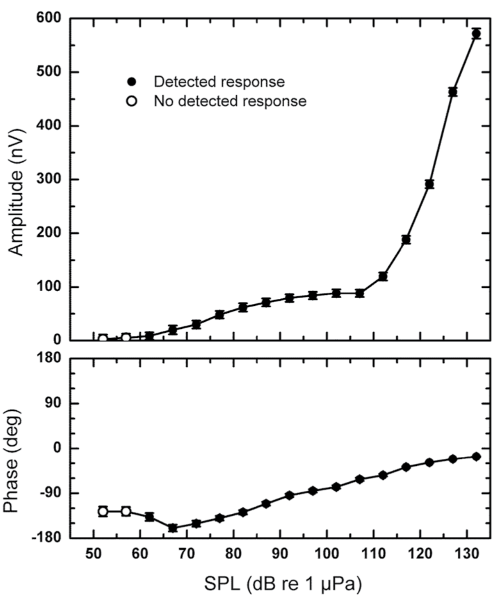

Figure 6. ASSR amplitude and phase are plotted as functions of stimulus SPL for 30 kHz f_c AM/FM stimuli. Symbols indicate mean values (± 95% CI) for 1,000 epochs.

Examples of the MSC approach are shown in Figure 9 for SPLs of 52, 62, 82, and 102 dB re 1 µPa (note the differences in vertical and horizontal scales). The complex amplitudes at f_{AM} from the grand average (filled circle) and from each of the *Q* = 20 subaverages (+) are plotted as vectors using the real and imaginary parts in Cartesian coordinates. Tight clustering of the subaverage vectors about the grand average vector indicates close amplitude and phase relationships between subaverages and will lead to relatively high MSC and detected responses. Large scatter in the subaverage vectors reveals random phase behavior and will lead to low MSC and no detected responses. A visual comparison may be made by comparing the grand average vector to the circle centered at the origin of each graph. This circle has a radius $r = \sqrt{MSC_{crit} \cdot MP}$, where *MP* is the mean power (the average of the subaverage powers), so if the grand average vector lies outside this circle (i.e., the grand average magnitude is greater than the circle radius), then MSC > MSC_{crit} and a response is detected. MSC values calculated for the entire data set (including the data of Figure 9) are shown in Figure 10 as a function of stimulus SPL. As with the *F* ratio, the MSC dips slightly between 100 to 120 dB re 1 µPa, presumably due to an increase in the noise. Responses were detected at SPLs ≥ 62 dB re 1 µPa (*Q* = 20, α = 0.01).

AEP and behavioral thresholds (Figure 11) showed significant high frequency hearing loss as previously observed for BLU (Finneran & Houser, 2006). Agreement between the AEP and behavioral thresholds was good at frequencies between 30 to 60 kHz (differences within 3 dB). Differences between AEP and behavioral thresholds were 16, 11, and 15 dB at 10, 20, and 80 kHz, respectively.

Discussion

Effective calibration of SAM tones in a small reverberant volume can be difficult because interference between the direct and reflected waves will affect the AM depth as well as the received SPL. Since the amplitude of the ASSR depends, in part, on modulation depth (Supin & Popov, 1995), fluctuations in

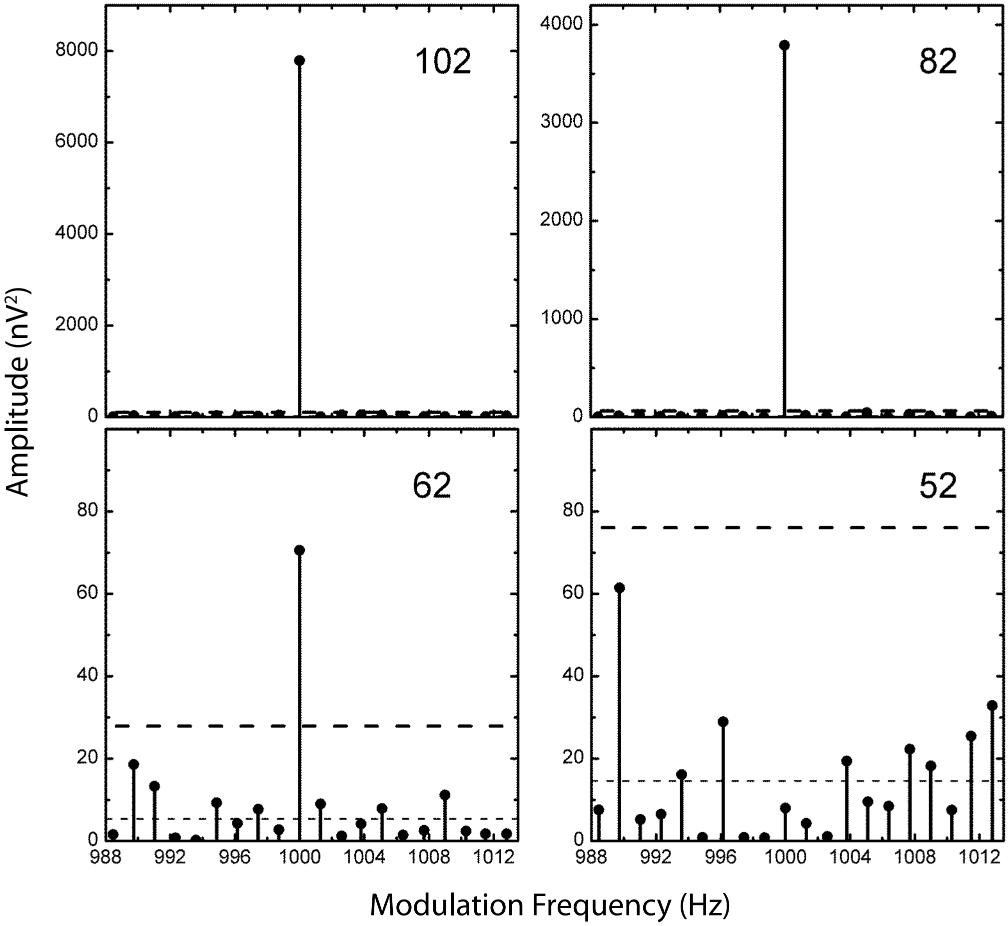

Figure 7. Example frequency spectra for significance testing using the *F* ratio; the number in each panel refers to the stimulus SPL in dB re 1 µPa. The thin dashed line represents the noise power estimate (P_n), and the thick dashed line represents $F_{crit} \cdot P_n$.

AM depth can have a profound effect on the amplitude of the ASSR. Although combination AM/FM tones have been used previously to increase the amplitudes of evoked responses (Picton et al., 1987b; Cohen et al., 1991; Dimitrijevic et al., 2002), in the present study, AM/FM stimuli were used to help control the acoustic stimulus received by the subject. Calibration measurements (with and without the subject present) and the resulting ASSR waveforms/spectra indicated that the combination AM/FM stimuli were an effective means of generating steady-state evoked potentials while reducing the influence of the confined acoustic test space within the pool. Of course, the use of the AM/FM stimuli resulted in an increase in stimulus bandwidth compared to SAM tones. Comparisons between pure tone and FM behavioral thresholds (Finneran & Schlundt, 2006) indicate that a 10% FM bandwidth would produce only minor differences from pure-tone thresholds (generally < 6 dB).

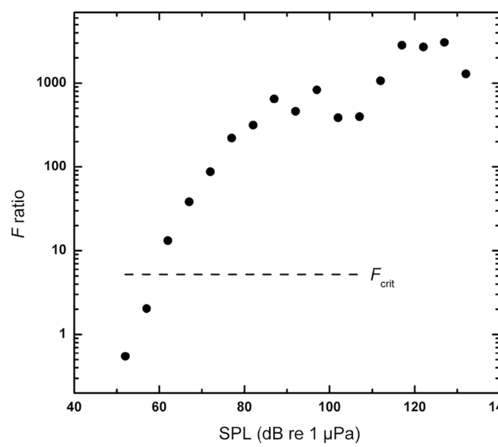

Figure 8. *F* ratio (m = 19, α = 0.01) calculated for the ASSR waveforms and spectra of Figures 4 & 5 as a function of stimulus SPL; responses were detected at SPLs ≥ 62 dB re 1 µPa.

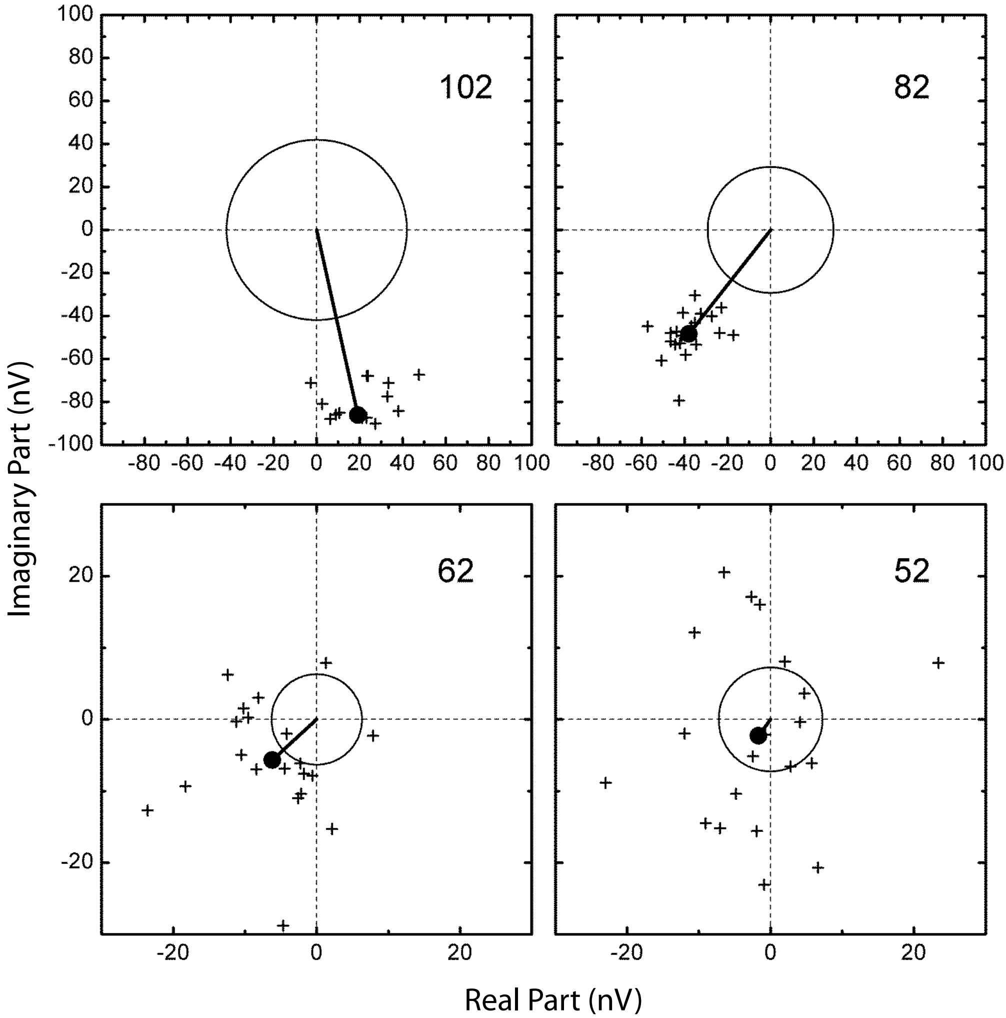

Figure 9. MSC calculation examples for 30 kHz AM/FM stimuli at 52, 62, 82, and 102 dB re 1 μPa (indicated in upper right of each panel)

The *F* test and MSC produced equivalent results when applied to the 30 kHz data set. This is expected since Dobie & Wilson (1996) not only demonstrated that the performances of the techniques are equivalent but that the SNRs required for detection are identical when $m = Q - 1$ as in the present study ($m = 19$ adjacent frequencies; $Q = 20$ subaverages). The decision to use one technique or the other becomes one of personal preference, though it is probably influenced by the manner in which the data are collected. The *F* test does not require data segmentation but does require relatively long epochs to achieve the necessary SNR for detection. Long epochs not only require more time for collection but increase the "cost" associated with rejecting epochs because of large artifacts. Visual examination of power spectra provides clear visual feedback on the likelihood of a response detection using the *F* test. In contrast, MSC is more difficult to visualize. MSC requires segment analysis, but this allows the collection of shorter epochs, which is advantageous when noise is nonstationary and epochs containing large artifacts must be rejected. MSC is also more cleanly applied to the analysis of intermittent data obtained using repetitive short SAM tones since each subaverage has the same duration as a single epoch. For the *F* test, intermittent data, as well as data obtained before and after epochs rejected for large artifacts, must be appended in such a way as to not introduce discontinuities into the waveform before frequency analysis. For threshold testing in

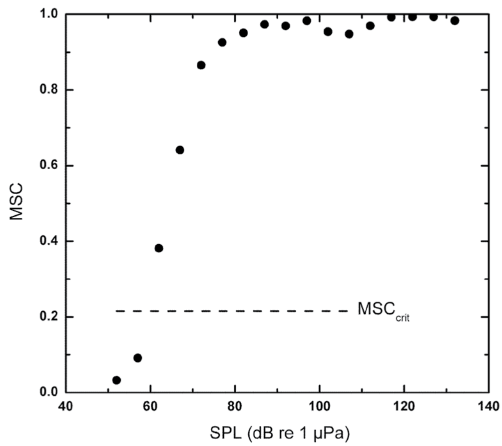

Figure 10. MSC ($Q = 20$, $\alpha = 0.01$) calculated for the ASSR waveforms and spectra of Figures 4 & 5 as a function of stimulus SPL; responses were detected at SPLs ≥ 62 dB re 1 μPa.

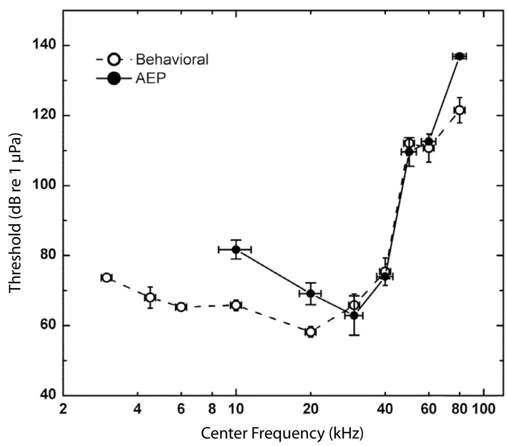

Figure 11. Behavioral and evoked potential audiograms; symbols and vertical error bars represent the mean ± 1 SD; The horizontal error bars indicate the frequency content of the stimuli.

the present study, MSC was chosen because the data collection method fit nicely with segment analysis.

Regardless of the specific technique, ORD methods offer substantial benefits to AEP testing. In addition to eliminating the effects of observer bias and experience and eliminating variability (within and between observers), ORD methods can also improve the speed and efficiency of data collection by providing a clear yes/no response for each measurement. Adaptive procedures can then be implemented to automatically adjust stimulus SPL from one measurement to the next, resulting in significant time savings. AEP methods are becoming increasingly applied to species and individuals to whom access and testing time are limited (Nachtigall et al., 2005; Cook et al., 2006; Mooney et al., 2006), so even modest improvements in testing time may be important. In the present study, 86% of the threshold estimates required seven or fewer individual AEP measurements, so thresholds were normally obtained within 4 to 5 min, even though all measurements began at 120 dB re 1 μPa, and the task was treated as if the subject's thresholds were unknown.

ORD techniques also allow alternative threshold definitions to be used. For example, threshold could be defined by interpolating within the MSC or *F* ratio data as a function of SPL to find the stimulus SPL corresponding to the critical value —the SPL producing a "just detectable" response. If threshold were defined in this manner, one could use a cost function such as $E = |MSC(SPL) - MSC_{crit}|$ and apply standard techniques to minimize the cost function and optimally reach threshold (Adby & Dempster, 1974).

Agreement between the ASSR thresholds and behavioral thresholds was excellent between 30 to 60 kHz. At 10, 20, and 80 kHz, the AEP thresholds were higher than the behavioral thresholds, though still within the range of differences between AEP and behavioral thresholds commonly cited (e.g., Rance et al., 1995; Vander Werff & Brown, 2005). Comparison between behavioral and ASSR thresholds have often exhibited increasing differences at the lower and higher frequencies (Yuen et al., 2005; Finneran & Houser, 2006), where it seems that the AM stimuli are not as effective in generating ASSRs. The effective received level for the lower frequency AM/FM tones used in the present study may also have been overestimated somewhat if the stimuli spectra were too large to remain within a single auditory filter; however, ASSR amplitudes typically decrease with increasing stimulus bandwidth, resulting in lower thresholds.

Acknowledgments

The authors thank Linda Green, Laura Lewis, and Randall Dear for animal training and technical assistance. Financial support was provided by the U.S. Office of Naval Research Marine Mammal Science and Technology Program.

Literature Cited

Adby, P. R., & Dempster, M. A. H. (1974). *Introduction to optimization techniques.* London: Chapman and Hall. 204 pp.

Amos, D. E., & Koopmans, L. H. (1963). *Tables of the distribution of the coefficient of coherence for stationary* *bivariate gaussian processes.* Livermore, CA: Sandia Corporation. 328 pp.

Bendat, J. S., & Piersol, A. G. (1986). *Random data* (2nd ed.). New York: Wiley. 566 pp.

Brillinger, D. R. (1978). A note on the estimation of evoked response. *Biological Cybernetics*, *31*(141), 141-144.

Campbell, F. W., Atkinson, J., Francis, M. R., & Green, D. M. (1977). Estimation of auditory thresholds using evoked potentials. In J. E. Desmedt (Ed.), *Progress in clinical neurophysiology* (pp. 68-78). Basel, Switzerland: Karger.

Cohen, L. T., Rickards, F. W., & Clark, G. M. (1991). A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans. *Journal of the Acoustical Society of America*, *90*(5), 2467-2479.

Cook, M. L. H., Varela, R. A., Goldstein, J. D., McCulloch, S. D., Bossart, G. D., Finneran, J. J., et al. (2006). Beaked whale auditory evoked potential hearing measurements. *Journal of Comparative Physiology A*, *192*, 489-495.

Cornsweet, T. N. (1962). The staircase method in psychophysics. *American Journal of Psychology*, *75*, 485-491.

Dimitrijevic, A., John, M. S., Van Roon, P., Purcell, D. W., Adamonis, J., Ostroff, J., et al. (2002). Estimating the audiogram using multiple auditory steady-state responses. *Journal of the American Academy of Audiology*, *13*, 205-224.

Dobie, R. A., & Wilson, M. J. (1989). Analysis of auditory evoked potentials by magnitude-squared coherence. *Ear and Hearing*, *10*, 2-13.

Dobie, R. A., & Wilson, M. J. (1993). Objective response detection in the frequency domain. *Electroencephalography and Clinical Neurophysiology*, *88*, 516-524.

Dobie, R. A., & Wilson, M. J. (1994a). Phase weighting: A method to improve objective detection of steady-state evoked potentials. *Hearing Research*, *79*(1-2), 94-98.

Dobie, R. A., & Wilson, M. J. (1994b). Objective detection of 40 Hz auditory evoked potentials: Phase coherence vs. magnitude-squared coherence. *Electroencephalography and Clinical Neurophysiology*, *92*(5), 405-413.

Dobie, R. A., & Wilson, A. J. (1995). Objective versus human observer detection of 40-Hz auditory-evoked potentials. *Journal of the Acoustical Society of America*, *97*(5), 3042-3050.

Dobie, R. A., & Wilson, M. J. (1996). A comparison of *t* test, *F* test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids. *Journal of the Acoustical Society of America*, *100*(4), 2236-2246.

Dolphin, W. F. (1995). Steady-state auditory evoked potentials in three cetacean species elicited using amplitude-modulated stimuli. In R. A. Kastelein, J. A. Thomas, & P. E. Nachtigall (Eds.), *Sensory systems of aquatic mammals* (pp. 25-47). Woerden, The Netherlands: De Spil.

Dolphin, W. F. (2000). Electrophysiological measures of auditory processing in odontocetes. In W. W. L. Au, A. N. Popper, & R. R. Fay (Eds.), *Hearing by whales and dolphins* (pp. 294-329). New York: Springer-Verlag.

Dolphin, W. F., & Mountain, D. C. (1992). The envelope following response: Scalp potentials elicited in the mongolian gerbil using sinusoidally AM acoustic signals. *Hearing Research*, *58*, 70-78.

Dolphin, W. F., Au, W. W. L., Nachtigall, P. E., & Pawloski, J. (1995). Modulation rate transfer functions to low-frequency carriers in three species of cetaceans. *Journal of Comparative Physiology A*, *177*, 235-245.

Egan, J. P., Greenberg, G. Z., & Schulman, A. I. (1961). Operating characteristics, signal detectability, and the method of free response. *Journal of the Acoustical Society of America*, *33*, 993-1007.

Finneran, J. J., & Houser, D. S. (2006). Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*). *Journal of the Acoustical Society of America*, *119*(5), 3181-3192.

Finneran, J. J., & Houser, D. S. (2007, in press). Bottlenose dolphin (*Tursiops truncatus*) steady-state evoked responses to multiple simultaneous sinusoidal amplitude modulated tones. *Journal of the Acoustical Society of America*, *121*(3).

Finneran, J. J., & Schlundt, C. E. (2006). Acoustic field measurements and bottlenose dolphin hearing thresholds using single-frequency and frequency-modulated tones. *Journal of the Acoustical Society of America*, *120*, 3227(A).

Finneran, J. J., Carder, D. A., Schlundt, C. E., & Ridgway, S. H. (2005). Temporary threshold shift (TTS) in bottlenose dolphins (*Tursiops truncatus*) exposed to mid-frequency tones. *Journal of the Acoustical Society of America*, *118*(4), 2696-2705.

Finneran, J. J., Schlundt, C. E., Carder, D. A., & Ridgway, S. H. (2002). Auditory filter shapes for the bottlenose dolphin (*Tursiops truncatus*) and the white whale (*Delphinapterus leucas*) derived with notched noise. *Journal of the Acoustical Society of America*, *112*(1), 322-328.

Gans, D., Del Zotto, D., & Gans, K. D. (1992). Bias in scoring auditory brainstem responses. *British Journal of Audiology*, *26*, 363-368.

Hall, J. W. (1979). Auditory brainstem frequency following responses to waveform envelope periodicity. *Science*, *205*, 1297-1299.

Hotelling, H. (1931). The generalization of student's ratio. *Annals of Mathematical Statistics*, *2*, 360-378.

Houser, D. S., & Finneran, J. J. (2006a). Variation in the hearing sensitivity of a dolphin population obtained through the use of evoked potential audiometry. *Journal of the Acoustical Society of America*, *120*(6), 4090-4099.

Houser, D. S., & Finneran, J. J. (2006b). A comparison of underwater hearing sensitivity in bottlenose dolphins (*Tursiops truncatus*) determined by electrophysiological and behavioral methods. *Journal of the Acoustical Society of America*, *120*(3), 1713-1722.

Houser, D. S., Crocker, D. E., Kastak, C., Mulsow, J., & Finneran, J. J. (2007, this issue). Auditory evoked potentials in northern elephant seals (*Mirounga angustirostris*). *Aquatic Mammals*, *33*(1), 110-121.

Jerger, J., Chmiel, R., Frost, J. D., Jr., & Coker, N. (1986). Effects of sleep on the auditory steady state evoked potential. *Electrophysiologic Techniques in Audiology and Otology*, *7*(4), 240-245.

Kastelein, R. A., Bunskoek, P., & Hagedoorn, M. (2002). Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals. *Journal of the Acoustical Society of America*, *112*(1), 334-344.

Kuwada, S., Batra, R., & Maher, V. L. (1986). Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hearing Research*, *21*, 179-192.

Lemonds, D. W., Au, W. W. L., Nachtigall, P. E., Vlanchos, S., & Roitblat, H. L. (1997). Auditory frequency selectivity and masked hearing capabilities in an Atlantic bottlenose dolphin. *Journal of the Acoustical Society of America*, *102*(5), 3102(A).

Lins, O. G., Picton, P. E., Picton, T. W., Champagne, S. C., & Durieux-Smith, A. (1995). Auditory steady-state responses to tones amplitude-modulated at 80-110 Hz. *Journal of the Acoustical Society of America*, *97*(5), 3051-3063.

Lins, O. G., Picton, T. W., Boucher, B. L., Durieux-Smith, A., Champagne, S. C., Moran, L. M., et al. (1996). Frequency-specific audiometry using steady-state responses. *Ear and Hearing*, *17*, 81-96.

Mooney, T. A., Nachtigall, P. E., & Yuen, M. M. L. (2006). Temporal resolution of the Risso's dolphin, *Grampus griseus*, auditory system. *Journal of Comparative Physiology A*, *192*, 373-380.

Nachtigall, P. E., Supin, A. Ya., Pawloski, J., & Au, W. W. L. (2004). Temporary threshold shifts after noise exposure in the bottlenose dolphin (*Tursiops truncatus*) measured using evoked auditory potentials. *Marine Mammal Science*, *20*(4), 673-687.

Nachtigall, P. E., Yuen, M. M. L., Mooney, T. A., & Taylor, K. A. (2005). Hearing measurements from a stranded infant Risso's dolphin, *Grampus griseus*. *The Journal of Experimental Biology*, *208*, 4181-4188.

Picton, T. W., Vajsar, J., Rodriguez, R., & Campbell, K. B. (1987a). Reliability estimates for steady-state evoked potentials. *Electroencephalography and Clinical Neurophysiology*, *68*(2), 119-131.

Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J., & Maiste, A. C. (1987b). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *Journal of the Acoustical Society of America*, *82*(1), 165-178.

Popov, V. V., & Supin, A. Ya. (1990). Auditory brainstem responses in characterization of dolphin hearing. *Journal of Comparative Physiology A*, *166*(3), 385-393.

Rance, G., Rickards, F. W., Cohen, L. T., De Vidi, S., & Clark, G. M. (1995). The automated prediction of hearing thresholds in sleeping subjects using auditory steady-state evoked potentials. *Ear and Hearing*, *16*(5), 499-507.

Rickards, F. W., & Clark, G. M. (1984). Steady state evoked potentials to amplitude modulated tones. In R. H. Nodar & C. Barber (Eds.), *Evoked potentials II* (pp. 163-168). Boston: Butterworth.

Rose, D. E., Keating, L. W., Hedgecock, L. D., Schreurs, K. K., & Miller, K. E. (1971). Aspects of acoustically evoked responses: Inter-judge and intra-judge reliability. *Archives of Otolaryngology*, *94*, 347-350.

Stapells, D. R., Makeig, S., & Galambos, R. (1987). Auditory steady-state responses: Threshold prediction using phase coherence. *Electroencephalography and Clinical Neurophysiology*, *67*(3), 260-270.

Stapells, D. R., Linden, D., Suffield, J. B., Hamel, G., & Picton, T. W. (1984). Human auditory steady state potentials. *Ear and Hearing*, *5*(2), 105-113.

Supin, A. Ya., & Popov, V. V. (1995). Envelope-following response and modulation transfer function in the dolphin's auditory system. *Hearing Research*, *92*, 38-46.

Supin, A. Ya., Popov, V. V., & Mass, A. M. (2001). *The sensory physiology of aquatic mammals.* Boston: Kluwer Academic Publishers. 332 pp.

Tucci, D. L., Wilson, M. J., & Dobie, R. A. (1990). Coherence analysis of scalp responses to amplitude-modulated tones. *Acta Oto-Laryngologica*, *109*(3-4), 195-201.

Vander Werff, K. R., & Brown, C. J. (2005). Effect of audiometric configuration on threshold and suprathreshold auditory steady-state responses. *Ear and Hearing*, *26*, 310-326.

Victor, J. D., & Mast, J. (1991). A new statistic for steady-state evoked potentials. *Electroencephalography and Clinical Neurophysiology*, *78*(5), 378-388.

Yuen, M. M. L., Nachtigall, P. E., Breese, M., & Supin, A. Ya. (2005). Behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*). *Journal of the Acoustical Society of America*, *118*(4), 2688-2695.

Zurek, P. M. (1992). Detectability of transient and sinusoidal otoacoustic emissions. *Ear and Hearing*, *13*(5), 307-310.