***Mystic Responses - Acoustic Study***

**File No. 22629 Comments received during the public comment period regarding Study 3—Hearing and physiological response to anthropogenic sound sent to Mystic Aquarium January 3, 2020. Mystic Aquarium's responses received January 24, 2020. NMFS' follow-up questions were sent on February 13, 2020, and responses were received February 25, 2020. Additional follow up questions were sent on March 2, 2020, and final responses were received March 20, 2020.**

**Comment 1(a):** What is meant by "masked hearing thresholds" and how does this differ from a temporary threshold shift?

**Mystic Response 1(a):** As defined by Hawkins and Chapman (Hawkins and Chapman 1975) 'Masked hearing thresholds' are hearing thresholds 'When the detection of one sound is impaired in the presence of another'. Trickey, Bransetter and Finneran (Trickey et al. 2010) similarly note 'Masking occurs when noise interferes with or obscures acoustic signals and thereby hinders the detection, discrimination, or recognition abilities. Johnson first studied this phenomenon in cetaceans in 1968 (Johnson, 1968). Temporary threshold shift is the temporary loss of hearing (increase in threshold) typically noted after a noise exposure. These are extensive hearing topics and are well-defined in Southall et al (Southhall et al. 2008) and reviewed by Finneran and Branstetter (Finneran and Branstetter 2013).

**Comment 1(b):** Specifically, are they planning on measuring hearing during noise exposure or afterwards? There is conflicting information. On page 6, it says "masked hearing collected while noise is projected," but from page 35 of the application, it says "Noise-related AEP hearing tests will follow each noise exposure session." It is unclear how this differs from a temporary threshold shift (TTS) study, if it is truly happening afterwards, because it looks like they are expecting some shift in hearing (although, they are playing sounds at much lower levels than would cause a temporary threshold shift, which is defined as a 6 dB increase from threshold) and measuring it after the noise exposure (rather than during, which is more consistent with masked hearing thresholds).

**Mystic Response 1(b):** We are proposing masking hearing experiments thus measuring hearing during the noise exposure. However, it's important to measure hearing before and after to ensure that TTS does not occur and to determine the amount of masking that did occur during the noise exposure. So this information is not conflicting, it's complimentary. We are doing both. This differs from a TTS study in (a) the level of noise projected and (b) typically in a TTS study, hearing is not measured during the noise exposure. Also note, TTS and masking both result in increased thresholds, but typically differ in the auditory mechanisms that drive that increase. But in masking, once the noise is off, hearing should be at the 'normal' or baseline levels. There is no recovery time in masking as there is in TTS.

**NMFS Follow-up to Response Comment 1(b) Part 1:** The second sentence regarding when hearing will be measured is helpful information and should be included in the application to be more clear what is being measured and why.

**Mystic Response to Follow-up 1(b) Part 1**: We will add the highlighted information to the application.

**NMFS Follow-up to Response Comment 1(b) Part 2:** Regarding the statement that "there is no recovery time in masking as there is in TTS." This is why it is important to measure afterwards to ensure TTS has not occurred.

**Mystic Response to Follow-up 1(b) Part 2**: Correct, we are measuring hearing after exposures to monitor for potential TTS.

**Comment 1(c):** In the application, they mention a proof concept study where it indicated "masking in beluga sensitivity exceeding 20 dB from baseline hearing," which seems to indicate a 20-dB threshold shift (again, it is unclear whether they measured this shift during or after exposure).

**Mystic Response 1(c):** Yes, we will be measuring for a ca. 20 dB increase in thresholds (due to masking) during the exposure, but not afterwards.

**NMFS Follow-up to Response Comment 1(c):** Do you know the maximum shift in hearing during masking that would not result in subsequent TTS? Do you have data you can reference? So immediately after the sound stopped, hearing thresholds were at baseline? It appears that the proof of concept study showed at least in one situation a measured 20-dB shift with no subsequent TTS.

**Mystic Response to Follow-up 1(c):** No, this is why we will use the user spreadsheet tool (NMFS 2018) to define a maximum SELcum based on the noise source characteristics and duration of exposure. The spreadsheet results will guide us to make sure we will never reach TTS. Furthermore, we will measure hearing just after the end of the exposure trial to make sure there was no TTS induced in any of the trials.

**Comment 1(d):** More detail is needed on how what they are doing does or doesn't relate to TTS, regardless of whether they are playing sound at level that is the TTS onset threshold in the Technical Guidance because again, they are expecting some sort of shift in hearing sensitivity.

**Mystic Response 1(d):** See points explained above. The experiments proposed here do not imply TTS. The increase in hearing thresholds are caused by noise masking the ability to hear sounds below the dB levels of the noise. There is no physiological impact causing a change in hearing thresholds (i.e., TTS). Thus the work proposed is likely generating differences of 20 dB in hearing from the baseline levels (i.e., hearing threshold in normal ambient noise conditions) and the masked levels (i.e., hearing thresholds increased because the projected noise is impeding the whale to hear at lower dB levels). It should be clear this is not a threshold shift of 20 dB caused by TTS.

**NMFS Follow-up to Response Comment 1(d):** Is a 20dB difference in hearing the max expected? What happens if you measure over 30 dB? Will you alter the protocol?

**Mystic Response to Follow-up 1(d):** We will not alter the protocol unless we detect TTS. We cannot anticipate how much masking is generated by the noise playback trials. But we can anticipate, using the user spreadsheet tool (NMFS 2018), what exposure is expected to trigger TTS, which we will make sure is never reached in any exposure trial. In the unlikely event of detecting TTS at any frequency tested just after an exposure trial, we will stop the study, reassess what was the actual SELcum reached for that whale that day, and define how much we need to increase our buffer to avoid further TTS. We are currently proposing to set a buffer of 10 dB below the onset of TTS identified in the user spreadsheet tool (NMFS 2018).

**NMFS Follow-up regarding TTS:** We should all (applicant, NMFS) make sure we are on the same page as what constitutes TTS onset. In the NMFS Technical Guidance, we define this as a 6 dB threshold shift from baseline hearing. We assume the applicant is defining this the same way, but it would be good to confirm.

**Mystic Response to Follow-up regarding TTS:** Yes, we will follow the TTS definition from NMFS Technical guidance of at least 6 dB shift in threshold.

**NMFS Follow-up regarding Acoustical Guidance:** The applicant mentions using the 2018 optional User Spreadsheet tool to ensure their exposures do not exceed our Technical Guidance's TTS onset thresholds. It would be good to confirm that they have an "unlocked" version of this Spreadsheet, so they can change the thresholds in it from PTS to TTS (our public tool only is for PTS onset, not TTS onset). If it is unlocked, they could even set these thresholds 10 dB lower, which is the buffer they are proposing. If they need an unlocked version of this Spreadsheet, NMFS will send it.

**Mystic Response to Follow-up regarding Acoustical Guidance**: Yes, we received the latest version of the unlocked spreadsheet. We will replace the PTS thresholds by TTS -10 dB and will send it back for confirmation. (Note: Mystic submitted the spreadsheet on May 28, 2020.)

**Comment 2 (similar to Comment 1):** Mystic did not make clear whether it plans to conduct masked hearing threshold tests, in which AEPs are conducted *during* sound exposure[1], or actual threshold shift (i.e., temporary threshold shift [TTS]) tests, in which the amount of threshold shift is measured *after* sound exposure. In the project description section of the application, Mystic indicated that it would compare baseline and masked hearing thresholds determined *while* the sound is projected. However, in the methods section, it stated that sound-related AEP hearing tests *would follow* each noise exposure session. It was initially assumed that Mystic was proposing to conduct masked hearing threshold tests and had inadvertently included incorrect methods throughout that section. But, Mystic noted in the project description section of the application that the controlled sound exposure experiments, mentioned previously herein, that were conducted on the single beluga whale indicated masking in beluga hearing sensitivity, exceeding 20 dB from baseline hearing. That assertion indicates that a 20-dB threshold shift occurred, which is more indicative of an actual threshold shift test. Onset TTS is defined as a shift of at least 6 dB (Southall et al. 2007, NMFS 2018, Southall et al. 2019) and therefore a 20- dB shift is much greater than onset TTS[2]. In short, Mystic's objectives do not comport with the proposed methods to fulfill those objectives.

[1]Similar to methods used by Terhune and Ronald (1975).

[2]Permanent threshold shift (PTS) is defined as 40 dB or more of TTS.

**Mystic Response 2:** We are proposing masking hearing experiments thus measuring hearing during the noise exposure. However, it's important to measure hearing before and after to ensure that TTS does not occur and to determine the amount of masking that did occur during the noise exposure. So this information is not conflicting, it's complimentary. We are doing both. This differs from a TTS study in (a) the level of noise projected and (b) typically in a TTS study, hearing is not measured during the noise exposure. Also note, TTS and masking both result in increased thresholds, but typically differ in the auditory mechanisms that drive that increase. But in masking, once the noise is off, hearing should be at the 'normal' or baseline levels. There is no recovery time in masking as there is in TTS. And yes, we will be measuring for a ca. 20 dB increase in thresholds (due to masking) during the exposure, but not afterwards.

The experiments proposed here do not imply TTS. The increase in hearing thresholds is caused by noise masking the ability to hear sounds below the dB levels of the noise. There is no physiological impact causing a change in hearing thresholds (i.e., TTS). Thus the work proposed is likely generating differences of 20 dB in hearing from the baseline levels (i.e., hearing threshold in normal ambient noise conditions) and the masked levels (i.e., hearing thresholds increased because the projected noise is impeding the whale to hear a lower dB levels). It should be clear this is not a threshold shift of 20 dB caused by TTS.

**NMFS Follow-up to Response Comment 2, Part 1:** Maybe stating what level of noise WOULD qualify as TTS to help differentiate the masking study would help – the way it reads it seems like TTS.

**Mystic Response to Follow-up 2, Part 1:** This is exactly what the user spreadsheet tool (NMFS 2018) will provide. Based on the planned trial (playback noise type, source level, duration of exposure, number of exposures in trial), we will obtain a SELcum threshold for onset of TTS. We will subtract 10 dB to that threshold and that will be the exposure level we will ensure we do not exceed in the trial. This will be calculated for each individual trial.

**NMFS Follow-up to Response Comment 2, Part 2:** We understand that there will be an increase in hearing threshold measured during the playback of sound and after the sound stops, hearing thresholds will immediately return back to baseline. In other words, there will be no residual threshold shift after the noise stops. However, it needs to be very clear on how you are ensuring that this doesn't end up resulting in TTS. How long after the end of the masking noise exposure will you measure hearing to make sure there is no residual or delayed threshold shifts? Immediately?

**Mystic Response to Follow-up 2, Part 2:** Yes, immediately, as soon as the trainers allow us to test hearing thresholds once the exposure is finished. This will normally require that the whale is rewarded after the exposure, then requested to move back to the position in the station to allow for measuring hearing, so hearing will be measured approximately 1 minute after the end of the exposure trial.

**NMFS Follow-up to Response Comment 2, Part 3:** Please clarify what you are measuring afterwards.

**Mystic Response to Follow-up 2, Part 3:** We will measure hearing thresholds at the same frequencies measured before the exposure trial, to make sure they stay at the same levels as before the exposure trial, which means there is no TTS induced by the trial.

**Comment 3:** If some sort of shift is expected during testing, there needs to be a discussion of measuring and ensuring hearing recovery, so that these minor shifts are completely recovered before starting another noise exposure session and to ensure repeated low-level exposure doesn't result in higher levels of TTS.

**Mystic Response 3:** This is a masking study and, as noted in our application, we expect to be well below TTS levels. But we will measure thresholds during (and after) the noise part of the tests to help quantify the amount of masking. Repeated baseline measurements before and after noise exposure also help ensure accurate baseline auditory measurements; hearing thresholds do vary slightly from day to day. This has been shown in many mammals.

**NMFS Follow-up to Response Comment 3, Part 1:** Same question as above – how soon after?

**Mystic Response to Follow-up 3, Part 1:** Thresholds will be measured approximately one minute before the exposure trial and one minute post exposure to allow time for the trainers to reward the whale, and then request it to go back to the station position.

**NMFS Follow-up to Response Comment 3, Part 2:** Please cite and quantify. What do you consider hearing returning to baseline (i.e., less than a 6 dB shift)? What if it is a 5-dB shift, will you wait until it is closer to baseline before starting another exposure?

**Mystic Response to Follow-up 3, Part 2:** Returning to baseline (or a baseline hearing threshold) is considered within the normal variation, typically within 1 standard deviation of the mean baseline thresholds (as in Mooney *et al.*, 2009a; Mooney *et al.*, 2009b). This allows us to capture the natural variability of any measurement, rather than an arbitrary dB value. Post-hoc we will empirically test (usually with ANOVA) whether thresholds increased during the masked hearing tests, or subsequent to the masker (see work by Finneran *et al.*, 2005; Kastelein *et al.*, 2017).

**NMFS Follow-up regarding "Full Recovery":** It needs to be clear that everyone (applicant, NMFS) is on the same page as to what is meant by "full recovery." In their responses, the applicant mentions using the criteria of 1 standard deviation within the normal variation as determining whether a threshold has returned to baseline. Can the applicant give some approximation as to how many dB 1 standard deviation typically is, realizing that it varies among individuals? For example, 1 standard deviation would not be more than 6 dB, correct (6 dB is the definition of TTS onset)?

**Mystic Response to Follow-up regarding "Full Recovery":** Yes, 1 standard deviation is typically within 2 to 3 dB difference, but in any case post-exposure thresholds should never be at or higher than 6 dB from pre-exposure thresholds in order to qualify as full recovery.

**Comment 4:** The application does not include information on whether other non-target beluga whales or other species held in the same pools as the whales exposed during the playback sessions and/or how Mystic will ensure other animals are not exposed to the playback sounds (e.g., moving the other animals to separate pools). If non-target animals will be exposed to playback sounds, information needs to be included on how many animals may be exposed and to what degree, and what measures will be implemented to minimize impacts.

**Mystic Response 4:** Non-target animals will be gated or moved to separate pools (as is typically for masking studies in pinnipeds and tursiops. The acoustic power of the proposed underwater speakers (both source level and directionality of the projection driven by the size of the speakers and the wattage) is not strong enough to exceed the baseline background noise levels of the facilities in adjacent pools, where the rest of the belugas will be housed while the experiment is ongoing.

**NMFS Follow-up to Response Comment 4:** Please provide the ambient noise measurements to justify this statement.

**Mystic Response to Follow-up 4:** SPL of the background noise at the Mystic Aquarium beluga facilities station where previous hearing research has been completed is 102.12 dB re 1micro Pa (100 Hz to 50 kHz range). If the speaker is projecting noise towards this station, which is away from any connecting gate to other pools, the indirect noise of the projection will be largely attenuated before it propagates into the adjacent pools (it is not simple spherical or cylindrical spreading, as there is no direct path from the speaker to the gates, and the gates are a confined funneled space for noise to propagate through). It is true that we have never measured the received levels in adjacent pools, but because we do not expect to project noise at high amplitudes, and background noise levels are not low in these facilities, we do not expect the exposure to whales in adjacent pools to be a concern.

**Comment 5:** Mystic's objective to 'quantify the frequency range and dB magnitude resulting from the noise sources' is not standard terminology. It is assumed that Mystic intends to quantify the threshold shift in terms of both the amplitude and the frequency over which the shift occurs.

**Mystic Response 5:** Yes, we aim to quantify the frequencies which are masked (if they are masked), and the difference (in dB) between baseline thresholds and masked thresholds. Note that the frequency range of masking is often not measured (see Branstetter and Finneran, 2008; and Trickey et al., 2011) but we feel in these studies that it's a helpful metric to evaluate the potential range of masking.

**Comment 6:** The application does not specify how AEPs would be conducted in general, let alone as part of masked hearing threshold tests. Information is missing regarding:

(A) the frequency range and specific frequencies that would be tested (for baseline and either during/after);
(B) whether clicks (impulsive) would be used in addition to pips (non-impulsive);
(C) the total active sound transmission time (including specifying how many 20-sec sound bursts, that are comprised of alternating 20-msec modulated tones and 30-msec silent periods and defined as pips, or how many clicks (including the pulse duration) would be emitted during a single baseline AEP session);
(D) the timeframe over which Mystic would collect baseline AEP data; and
(E) whether a full audiogram or thresholds at only specific frequencies would be collected for the masked hearing threshold tests.

**Mystic Response 6:**

- A(1) Recording AEPs is a well-established procedure. We will conduct AEP hearing tests following Castellote et al., 2014; Mooney et al., 2018, (for belugas) and other AEP odontocete studies. We first need to establish baseline audiograms for these animals. This will include testing from 4-180 kHz. Likely frequencies will include: 4, 5.6, 8, 11.2, 16, 22.5, 32, 45, 54, 80, 100, 120, 128, 150 and 180 kHz. These are typically octave and half-octave increments. On occasion we may use finer increments, such as quarter-octaves, if there is a need to understand a particular part of an audiogram better.
- A(2). For the masking study we expect to conduct hearing measurements on the same increments as above, but we do not expect to go as high in frequency, thus this will be a limited frequency range. As masking typically occurs in critical bands (but we must confirm this for each anthropogenic noise) we do not expect to measure much higher than an octave above the maximum higher ⅓ band of the noise. This would be 22 kHz with respect to the commercial ship noise. Going just higher than this we expect the auditory measurements to not go higher than 32 kHz for all other anthropogenic sounds.
- B. Yes, clicks will be used in addition to pips. This is standard for audiometry to ensure animals hear and a response can be generated, because these pulses span a broad range of frequencies.
- C. There are no solid answers for C and D as the number of AEP trials and consequently sound-on times is dependent on the animal's hearing ability. If an animal only hears to 50 kHz, there will fewer trials compared to if an animal hears up to 150 kHz. The number of trials for a particular animal to acquire an audiogram (based on wild belugas that hear very, very well) was ca. 60-80. There are 1000 sweeps (tone pip or click series) per record (trial). That can change with certain methods of objective response detection used in the ANSI AEP hearing standards (ANSI 2018). On the high side, that would be ca. 80,000 sweeps per audiogram. Although all of these are at, near, or below threshold, in approximately 20% of them the animal cannot hear (and in that case the animal does not generate an AEP response). We estimate about 5 frequencies per masking study, so about 40 records and 40,000 tone pips for each set of hearing measurements. Again, about 20% would be inaudible. The rest are at or near the level of audibility (very near detection threshold).

**NMFS Follow-up to Response Comment 6(C):** If 5 frequencies is 40,000 pips (for masking), then how can testing the entire audiogram (up to 15 frequencies) be only 80,000 sweeps, unless what they are doing during masking is different than collecting an audiogram? Please provide more detail to clarify.

**Mystic Response to Follow-up 6(C):** We are not collecting a whole audiogram during masking. We expect masking to only occur at the frequencies (or within the critical ratios) for which the noise is located (Branstetter and Finneran, 2008; Trickey et al., 2010; Branstetter et al., 2013; Branstetter et al., 2017). And this noise proposed is low frequency, thus we will focus on low frequencies and a small part of the hearing range.

**Mystic Response to Comment 6 continued:**

- D. Not sure if we understand D correctly but overall timeframe of the AEP data collection would be intermittent over the length of the permit and study period (i.e. 5 yrs).
- E. See responses to A(2)

**Comment 7:** Mystic noted that it would use tones as masking sounds but then specified that recordings from commercial ships, dredging rigs, aircraft, outboard motors, and impact hammers would be used—none of those recordings are considered tones. The frequency range and source levels of those recordings were omitted from the application and need to be included.

**Mystic Response 7:** To clarify, tones will not be used as the maskers, but rather the probe (hearing test) tones. The spectra and received levels were noted in Castellote et al., 2018, but we list them here Table 1. The frequency range of the masking sounds (commercial ship noise, dredging noise, etc.) is presented below as the lower and higher ⅓ octave band containing the peak value for all the recordings planned to be used. This provides a sense of the frequency distribution of the acoustic energy of these signals that has the potential to mask hearing for communication purposes (note that detailed information on these signals is in Castellote et al. 2018).

**NMFS Follow-up to Response Comment 7, Part 1:** It is unclear what is meant by "containing the peak value for all recordings planned to be used"? Are the frequencies listed in Table 1 the range of frequencies produced by the noise source? Where is the peak in the higher or lower 1/3 octave band or did you provide the two octave bands (lower an upper) with the most sound? To be clearer, could you provide the range of frequencies produced by the source (upper/lower), and then indicate where the peak is, and then finally indicate highest frequency you are going to test hearing (could do this as a footnote)?

**Mystic Response to Follow-up 7, Part 1:** The table presents the 1/3 octave center frequency range (lowest and highest frequency) containing the peak energy of the signal to be played back. It represents the 1/3 octave range that has the highest potential for masking. This information comes from Figure 5 in Castellote et al., 2018:

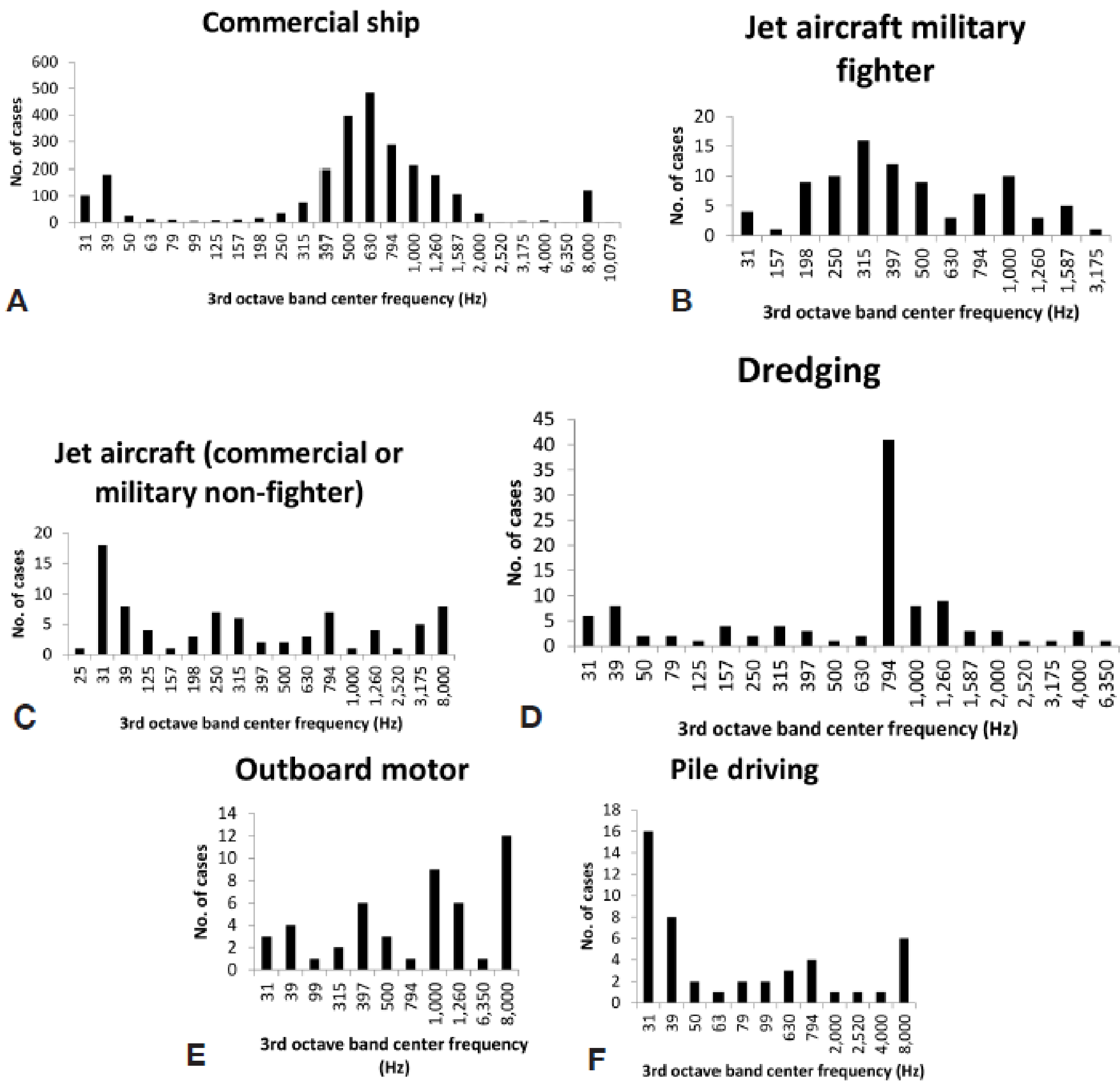


Figure 5.—Histograms with counts of 1/3 octave bands containing the maximum amplitude (peak band) per noise class. This histogram provides an overview of the dominant 1/3 octaves for each noise class identified in the sampled locations in Cook Inlet, Alaska. Noise classes included are, A) commercial ship, B) jet aircraft military fighter, C) jet aircraft (commercial or military non-fighter), D) dredging, E) outboard motor, and F) pile driving. For all other noise classes see Castellote et al. (2016c).

Based on this analysis, peak 1/3 octave band for each noise is:

| **Noise source** | **Peak band (1/3 octave) center frequency (Hz)** |
|---|---|
| **Commercial ship** | **630** |
| **Military fighter** | **315** |
| **Commercial aircraft** | **31** |
| **Dredging** | **794** |
| **Outboard** | **8000** |
| **Pile driving (impact hammer)** | **31** |

Highest frequency to be tested for hearing thresholds might be 128 or 150 kHz range. This depends on the amount of time we might have with the whale in station, which depends on the training abilities, whales' motivation, etc. and thus can't be defined with precision.

**Table 1. Range of 1/3 octave bands containing the maximum amplitude (peak band) per noise class proposed to be used as masking noise in the experiments, following Castellote et al. (2018).**

| Noise source | Lower 1/3 octave band | Higher 1/3 octave band |
|---|---|---|
| Commercial ship | 31 | 10,079 |
| Military fighter | 31 | 3175 |
| Commercial aircraft | 25 | 8000 |
| Dredging | 31 | 6350 |
| Outboard | 31 | 8000 |
| Pile driving (impact hammer) | 31 | 8000 |

**NMFS Follow-up to Response Comment 7, re: Table 1:** To clarify, is this the extent of sounds to be used for the masking study (e.g. are these just an example or are there more)?

**Mystic Response to Follow-up 7, re: Table 1:** This is the extend of sounds to be used for the masking study. The source level for these masking signals will depend on the design of each of the trials as we seek variation in the masking effects. But in any case, the speaker system factory specifications suggest it will not allow exceeding a maximum output level of 180 dB re 1 μPa @ 1m at 1 kHz, which for structured wideband noise is more likely to be in the ~170 dB.

**NMFS Follow-up to Response Comment 7, Part 2:** Please clarify that you will confirm that structured wideband noise is more likely to be in the ~170 dB.

**Mystic Response to Follow-up 7, Part 2:** Yes, field exposure calibrations will be completed before any exposure.

**NMFS Follow-up regarding Table 1:** Mystic's response to the original question about Table 1 was "The table presents the 1/3 octave center frequency range (lowest and highest frequency) containing the peak energy of the signal to be played back. It represents the 1/3 octave range that has the highest potential for masking. This information comes from Figure 5 in Castellote et al., 2018." From reading the response, it seems like the lower and higher bands provided in Table 1 represents the frequency range of peak energy for these signals. Also on page 4 of their responses to NMFS' original comments, they say "As masking typically occurs in critical bands (but we must confirm this for each anthropogenic noise) we do not expect to measure much higher than an octave above the maximum higher ⅓ band of the noise. This would be 22 kHz with respect to the commercial ship noise. Going just higher than this we expect the auditory measurements to not go higher than 32 kHz for all other anthropogenic sounds. "

Later on this same page they say "We estimate about 5 frequencies per masking study, so about 40 records and 40,000 tone pips for each set of hearing measurements"

Additional questions regarding Table 1:

(A) Based on sounds listed in Table 1, it is unclear how there would only be 5 frequencies tested for masking. Mystic lists 15 frequencies for the entire audiogram (4/5.6/8/11.2/16/22.5/32/45/54/80/100/120/128/150/180 kHz). Thus, if they are going up to 22 kHz for commercial ship noise (as indicated above), this is 6 frequencies and, for all other sounds up to 32 kHz (as indicated above) 7 frequencies would be included. Can the number of frequencies to be tested be clarified?

(B) Since all these sound sources produce sound down to 31 Hz, is there a concern that Mystic is only testing hearing down to 4 kHz? Plus, for military fighter, this source has its higher 1/3 octave band below 4 kHz (3.175 kHz), which is below what they are going to test. This needs to be addressed.

**Mystic Response regarding Table 1:** We will limit our threshold measurements to the frequency range where masking noise might have an effect in hearing (within an octave above/below the maximum/minimum higher ⅓ band of the noise). The speaker cannot play signals below 200 Hz or above 20 kHz, and our AEP system will not be able to get accurate thresholds below 4 kHz, thus we will focus on the range 4 kHz to 20 kHz. For masking signals where their peak energy is below the 4 kHz limit, we have no other option than measuring hearing at 4 kHz. Based on previous experience we believe we can collect 5 thresholds per trial, but this depends on the whale's motivation to follow trainers' instructions during each session. For the calculation of the maximum exposure (10 dB below TTS onset) per trial, will plan for 5 frequencies distributed across the frequencies of maximum received levels for each masking noise, within the frequency range specified above (4 kHz to 20 kHz). We might be able to get more frequencies if the trainers/veterinarian can keep the whale in station for longer periods. Therefore, in the best case scenario, we will not exceed 6 frequencies per trial.

| **Noise source** | **Lower 1/3 octave band** | **Higher 1/3 octave band** | **Frequencies to measure (predicted, tentative)** |
|---|---|---|---|
| Commercial ship | 31 | 10,079 | 4/5.6/8/11.2/16 |
| Military fighter | 31 | 3175 | 4/5.6/8/11.2/16 |
| Commercial aircraft | 25 | 8000 | 4/5.6/8/11.2/16 |
| Dredging | 31 | 6350 | 4/5.6/8/11.2/16 |
| Outboard | 31 | 8000 | 4/5.6/8/11.2/16 |
| Pile driving (impact hammer) | 31 | 8000 | 4/5.6/8/11.2/16 |

**Comment 8:** Mystic did not consider the sound emitted during the AEPs when it assessed its two scenarios (impact pile driving and ship noise) for the masked hearing threshold tests. Based on Mystic's scenario 1 for impact pile driving, it estimated that the weighted cumulative sound exposure level (SELcum) would be 158.3 dB re 1 μPa2-sec, which is less than the 160-dB re 1 μPa2-sec threshold[3]. However, Mystic did not specify whether it would conduct a baseline AEP before the first of two 3-minute playback sessions or whether it would ensure recovery prior to conducting the second playback session.

If those two sessions were to occur, a total of four AEP sessions would be conducted for this scenario—one baseline AEP before the first session, the AEP associated with session 1, the AEP to ensure recovery before conducting the second session, and the AEP associated with session 2. Sound emitted during those AEP sessions could exceed the threshold for scenario 1 (if both impulsive and non-impulsive sounds are emitted (e.g., impact pile driving sounds and AEP pips), the more conservative impulsive threshold is used to ensure that the threshold is not exceeded).

Further, it is unclear how Mystic could conduct up to three 15-minute sessions per day given that it proposed to emit sound at only 157.2 re 1 μPa2-sec at 1 m and for only two 3-minute sessions for scenario 1.

**Mystic Response 8:** The AEP acoustic stimuli is of very low SL, very short in duration (in the order of ms) and narrowband in nature, barely reaching audibility because the aim of the collection is to identify onset of hearing (hearing threshold); thus, it makes no sense to consider the AEP sound projection in the calculation of risk for TTS. Regarding the 15-minute sessions, the full procedure of positioning the belugas behaviorally on station and allowing breaks for positive reinforcement in between data collection periods, would last no more than 45 minutes. Within these 45 minutes, the whale would be positioned on station with the speaker actively playing noise only a small amount of time. As an example, we presented the 3-minute sessions with impact pile driving noise targeting a specific exposure. To clarify, scenarios 1 and 2 presented in the permit application are given as examples of how we will calculate the total exposure (SELcum) to ensure the amount of acoustic energy will never reach TTS thresholds as defined by NOAA (in fact, we propose to keep a 10 dB buffer to that limit as a precaution). We cannot pre-set a priori the exposure conditions of the sessions to be included in the experiment because we cannot anticipate the level of masking caused by the different playbacks (if we knew, this experiment would be

[3] Which is based on maintaining Mystic's proposed buffer of 10-dB less than the weighted TTS threshold.

useless). Thus, the experiment proposed requires to have flexibility in the source levels and duration of exposures (sessions). We have set limits in that flexibility to ensure that we will never approach TTS exposures.

**NMFS Follow-up to Response Comment 8, Part 1:** It would be helpful to provide some additional information on what the expected threshold of hearing is as a reminder of the low levels you are talking about. We agree that if the aim is to collect onset of hearing, then there shouldn't be a concern with TTS (e.g., if they were playing the sound loud enough to cause TTS, then this would interfere with collecting baseline line hearing measurements). You could perhaps talk about "effective quiet" to support that levels you are using to test onset of hearing would be way below this. The only place where there could be concern is after the masking noise, if you have to play the sound louder than before the noise in order to determine onset of hearing, it would mean the animal has TTS (which would be the larger concern). It then becomes a bit of a circular argument…if measuring onset of hearing causes hearing loss, then how do you ever measure recovery? How do you ever know what baseline hearing is?

**Mystic Response to Follow-up Comment 8, Part 1:** You know baseline hearing threshold because you've measured hearing repeatedly at a frequency before any experimental trials. You also have 'control' trials in which a no-noise signal (a sham) is 'played' to the animal. In actuality, no masker is actually played in these sessions. You can compare masked hearing thresholds to the control and baseline thresholds statistically (via ANOVA or similar) and in real time, by comparing to the natural standard deviation. Additionally, you are never playing the hearing test tones near the levels which may induce TTS. Masked thresholds are typically only increased by a few dB (Branstetter and Finneran, 2008; Trickey *et al.*, 2010; Branstetter *et al.*, 2013; Branstetter *et al.*, 2017), thus all hearing test tones may increase that amount, but do not need to go greater. For example, hearing test tones for belugas may be 30-80 dB rms, and as noted above, very short in duration, thus they carry very little energy (See Castellote et al., Mooney et al., 2018). TTS levels may often by near 160 dB rms for much longer durations to reach 180 dB SEL, 5 orders of magnitude higher. So we will never get in that cycle.

**NMFS Follow-up to Response Comment 8, Part 2:** To make the case that sound used to measure hearing before and after isn't accumulated and that the levels you are playing to measure hearing are below what is expected to cause TTS (e.g., even below effective quiet), providing some levels or max levels associated with their onset of the hearing procedures would put this in context. Those conducting the hearing aspect of this study (Castellote[4] and Mooney) have extensive experience measuring audiograms, especially with beluga whales. Thus, they have the experience to know if something is amiss during a baseline hearing measurement as to not keep pushing the received level higher and higher, close to inducing TTS. They can perhaps use their experience to bolster the response.

**Mystic Response to Follow-up Comment 8, Part 2:** See above; we expect animals to have sensitive hearing, near 40-80 dB rms (Castellote et al 2014, and Mooney et al., 2018) and TTS levels to be several orders of magnitude higher and for much longer durations (160-200 dB rms) (reported in all odontocete TTS papers). We don't know of any studies on effective quiet in odontocetes so we cannot empirically address this.

---

[4]Dr. Castellote of the NMFS Alaska Fisheries Science Center, Marine Mammal Laboratory did not participate in NMFS' review of the application.

**Comment 9:** There are few minor errors associated with the two example playback scenarios in the application:

(A) It says "Two playback session scenarios are presented below, as an example of the application of TTS **-20** dB threshold limit..." The "-20 dB" is a remnant from an earlier version of the application. This most recent version indicates, it should be **-10** dB below TTS (not 20).
(B) For their second example, they indicate that 765 x 2 is **1630** seconds, but it is really **1530** seconds (overall their calculation for determining unweighted SELcum is correct; they just made an error on this one step).

**Mystic Response 9:** Agreed. The proposed changes above are correct.

**Comment 10:** Mystic indicated in the project description section of the application that it planned to quantify directional hearing abilities of beluga whales from three to five different angles. However, in the methods section of the application, Mystic indicated that anthropogenic noise would be projected at 0, 90, and 180 degrees from an animal's head. Directional hearing tests are conducted using AEP tones (similar to methods used by Popov and Supin [2009]), not anthropogenic sound recordings from ships, dredging rigs, and impact hammers. If Mystic intended to determine how masking affects a beluga's directional hearing, then that should have been specified. At present, Mystic's stated objectives likely would not be achieved with the methods proposed. The directional hearing ability objective appears was mentioned only as an objective in the project description section of the application and was not discussed further in the justification and summary of published findings portion of that same section. Numerous papers have been published on directional hearing in odontocetes (e.g., Au and Moore 1984, Supin and Popov 1993, Kastelein et al. 2005, Popov et al. 2006) and specifically on beluga whales (e.g., Klishin et al. 2000, Mooney et al. 2008, Popov and Supin 2009). In addition, Mystic included only a single sentence describing its directional hearing methods in the application, which is insufficient.

**Mystic Response 10:** The rationale behind the concept of quantifying directionality in hearing in this study is to address how masking release could be achieved by a beluga, simply by turning its head away from the noise. We did not provide details on why the interest in projecting noise at different angles, since it is much simpler to devise a study in which a speaker is moved than to train a beluga to turn its head on command. By projecting noise at 0, 90 and 180 degrees from the longitudinal axis of the beluga we are simulating a beluga turning its head (or its body in the case of 180 degrees) away from the noise source comparing the masked hearing thresholds obtained from 0 degrees (on axis) sessions, to the 90 and 180 degrees sessions of same exposures (sessions of identical characteristics except for angle of noise projection) will inform how much masking is avoided by turning the head away.

**Comment 11:** Conducting hearing-related tests that have the potential to injure or harm an animal, even when a trained animal participates on a voluntary basis, should be considered intrusive research and be covered under a permit (for the beluga whales held under public display at Mystic Aquarium and not included on the permit application).

NMFS's implementing regulations define intrusive research as a procedure conducted for bona fide scientific research involving: A break in or cutting of the skin or equivalent, insertion of an instrument or material into an orifice, introduction of a substance or object into the animal's immediate environment that is likely either to be ingested or to contact and directly affect animal tissues (i.e., chemical substances), **or a stimulus directed at animals that may involve a risk to health or welfare or that may have an impact on normal function or behavior (i.e., audio broadcasts directed at animals that may affect behavior).**

**For captive animals (held for public display), this definition does not include procedures that** (1) are conducted by the professional staff of the holding facility or an attending veterinarian for purposes of animal husbandry, care, maintenance, or treatment, or a routine medical procedure that, in the reasonable judgment of the attending veterinarian, would not constitute a risk to the health or welfare of the captive animal or (2) **involve** either the introduction of a substance or object (i.e., as described in this definition) or **a stimulus directed at animals that, in the reasonable judgment of the attending veterinarian, would not involve a risk to the health or welfare of the captive animal** (50 C.F.R. § 216.3).

**Mystic Response 11:** This permit application is specifically aimed to cover the above. Our proposed study does not require reaching TTS, but the aim is to understand masking levels in hearing function occurring long before a source of noise approaches TTS levels. But because we need the flexibility to design exposure sessions of varied source levels, noise sources and durations, we are proposing an approach that will evaluate the exposure based on the characteristics of each session before these are executed, to confirm that TTS thresholds will never be reached. Whale behavior during these sessions will be constantly evaluated by the veterinarians and animal husbandry team at Mystic Aquarium, and playback and stationing will be terminated at any time and immediately, if requested. Thus, TTS threshold control is one form of safety. Ultimately, the behavior of the whales and the judgement of the staff will overdrive the execution of the experiment sessions.

**NMFS Follow-up to Response Comment 11:** Does this mean that the three whales currently on public display (not the subjects of the proposed permit) would also be included under the permit? Or, has the veterinarian determined that this would NOT rise the level of intrusive research for the public display animals (i.e., because the attending veterinarian has determined it would not be a risk to the health and welfare of the animals), and thus, a permit is not needed for the public display animals? If the veterinarian has determined this will not constitute a risk to the health and welfare of the animals, do you have a supporting statement you could provide?

**Mystic Response to Follow-up Comment 11:** The veterinary staff at Mystic Aquarium have determined that the proposed hearing-related studies do not pose undue risk to the health or welfare of the public display animals and therefore, an additional permit should not be required. Through review and assessment of the proposed study it was determined that since data acquisition does not require reaching TTS, the risk to the health and welfare of the current beluga population at Mystic Aquarium is negligible. In addition, the veterinary staff has inspected and reviewed all acoustic equipment for safety and deemed it to be low to no risk to the belugas. Through ongoing evaluations during previous acoustic sessions, veterinary staff and animal husbandry staff have not observed any adverse effects, however, if these were observed during the proposed study it would be immediately terminated to ensure the ongoing health and welfare of the animals.

**Comment 12:** Would activities cease if an animal was in distress or showed overt disturbance by the acoustic studies (or for any research procedure proposed)? If yes, how soon would the study again resume?

**Mystic Response 12:** Yes, activities would cease if the animal showed distress or overt disturbance by the acoustic studies (or for any research procedure proposed). The study would only resume after consultation and discussion with the SVP of Zoological Operations, Chief Clinical Veterinarian, Curator of Marine Mammals & Birds, VP of Research, Acoustician, and Trainer.

**References provided as part of Mystic's responses:**


ANSI (2018) Procedure for Determining Audiograms in Toothed Whales through Evoked Potential Methods. ANSI/ASA S3/SC16-2018 Melville, NY

Branstetter BK, Finneran JJ (2008) Comodulation masking release in bottlenose dolphins (*Tursiops truncatus*). J Acoust Soc Amer 124 (1):625-633

Branstetter, B. K., Trickey, J. S., Bakhtiari, K., Black, A., Aihara, H., and Finneran, J. J. (2013). "Auditory masking patterns in bottlenose dolphins (*Tursiops truncatus*) with natural, anthropogenic, and synthesized noise," J Acoust Soc Amer 133, 1811-1818.

Branstetter, B. K., Van Alstyne, K. R., Wu, T. A., Simmons, R. A., Curtis, L. D., and Xitco Jr, M. J. (2017). "Composite critical ratio functions for odontocete cetaceans," J Acoust Soc Amer 142, 1897-1900.

Castellote M, Mooney TA, Hobbs R, Quakenbush L, Goertz C, Gaglione E (2014) Baseline hearing abilities and variability in wild beluga whales (*Delphinapterus leucas*). J Exp Biol 217:1682-1691. doi:doi:10.1242/jeb.093252

Castellote M, Thayre B, Mahoney M, Mondragon J, Lammers MO, Small RJ (2018) Anthropogenic Noise and the Endangered Cook Inlet Beluga Whale, Delphinapterus leucas: Acoustic Considerations for Management. Marine Fisheries Review 80 (3):63-88

Finneran, J. J., Carder, D. A., Schlundt, C. E., and Ridgeway, S. H. (2005). "Temporary threshold shift in bottlenose dolphins (*Tursiops truncatus*) exposed to mid-frequency tones," J. Acoust. Soc. Am. 118, 2696–2705.

Finneran JJ, Branstetter BK (2013) Effects of noise on sound perception in marine mammals. In: Animal Communication and Noise. Springer, pp 273-308

Hawkins A, Chapman C (1975) Masked auditory thresholds in the cod, Gadus morhua L. Journal of comparative physiology 103 (2):209-226

Johnson, CS. Masked tonal thresholds in the bottlenosed porpoise. The Journal of the Acoustical Society of America 44, no. 4 (1968): 965-967.

Kastelein, R. A., Helder-Hoek, L., and Van de Voorde, S. (2017). "Effects of exposure to sonar playback sounds (3.5–4.1 kHz) on harbor porpoise (*Phocoena phocoena*) hearing," J Acoust Soc Amer 142, 1965-1975.

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.-(UK) 5, 565-567.

Mooney TA, Castellote M, Quakenbush L, Hobbs R, Gaglione E, Goertz5 C (2018) Variation in hearing within a wild population of beluga whales (*Delphinapterus leucas*). J Exp Biol 221:jeb171959. doi:doi: 10.1242/jeb.171959

Southall BL, Bowles AE, Ellison WT, Finneran JJ, Gentry RL, Greene Jr CR, Kastak D, Ketten DR, Miller JH, Nachtigall PE, Richardson WJ, Thomas JA, Tyack PL (2008) Marine mammal noise exposure criteria: Initial scientific recommendations. Aquat Mamm 33 (4). doi:DOI: 10.1578/AM.33.4.2007.411

Trickey JS, Branstetter BK, Finneran JJ (2010) Auditory masking of a 10 kHz tone with environmental, comodulated, and Gaussian noise in bottlenose dolphins (*Tursiops truncatus*). J Acoust Soc Amer 128 (6):3799-3804