



## Commentary

## Pseudoreplication of sound treatments in underwater exposure studies

Jeroen Hubert <sup>a, b, \*</sup>, Max Willems <sup>a, c</sup>, Judith M. Varkevisser <sup>a</sup><sup>a</sup> Institute of Biology Leiden, Leiden University, Leiden, the Netherlands<sup>b</sup> Marine Animal Ecology Group, Wageningen University, Wageningen, the Netherlands<sup>c</sup> Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany

## ARTICLE INFO

## Article history:

Received 30 January 2024

Initial acceptance 31 March 2024

Final acceptance 9 May 2024

Available online 31 July 2024

MS. number: 24-00075R

## Keywords:

anthropogenic noise

experimental design

pseudoreplication

sound exposure

sound playback

treatment design

underwater acoustics

underwater sound impact

Sound exposure studies require replicated sound treatments for the results to be representative for sound classes in general. Additionally, reused treatments in replicated designs need to be accounted for statistically. The lack hereof is referred to as simple and sacrificial pseudoreplication, respectively, and results should be interpreted accordingly. We quantified the occurrence of these issues and subsequent interpretation of results in 104 underwater sound exposure studies (2019–2023). The majority of the studies (85%) did not replicate sound treatments. From the ones that did, most did not statistically acknowledge the hierarchical structure of the data. Unreplicated treatment designs limit the generalizability of the findings. Nevertheless, only small differences were found in how the results of unreplicated and replicated treatment designs were interpreted. This commentary aims to provide guidance in the design, analysis and interpretation of sound exposure studies, which are equally valid for aquatic and terrestrial research.

© 2024 The Author(s). Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

The effects of sound on humans and nonhuman animals are increasingly gaining societal and scientific interest (Duarte et al., 2021; Kunc & Schmidt, 2019; WHO, 2011). It is estimated that 20% of the European Union population lives in areas where traffic noise levels are harmful to health (EEA, 2020). Increasing evidence shows that nonhuman animals are also impacted by noise pollution. Over the last 20 years, there is growing interest for research that focuses on the effects of underwater sound on aquatic animals (Duarte et al., 2021). This interest stems from the important role sound plays in communication and orientation of animals within the underwater environment, while this is also a noisy environment due to human activities. Numerous studies aim to gain insight into the effects of specific sound sources on aquatic animals and compare the effects of different sources or classes of sound. This provides input for impact assessments and mitigation measures and can yield insight into animal cognition. To draw valid conclusions from exposure experiments, sound treatments should be designed with care, and their effects should be analysed and interpreted accordingly.

It is widely accepted that animal experiments require sufficient sample sizes, in this case animal subjects, to draw more robust conclusions, control for confounding factors and enhance the generalizability of research findings. Animal subjects exhibit individual variability in terms of genetics, behaviour and physiology (MacKinlay & Shaw, 2023). Hence, individuals can exhibit unique characteristics or responses that may not be representative of the overall population. By replicating trials with a sufficiently sized sample of animals, results are more likely to provide meaningful insights into the broader population (Tipton et al., 2017).

Interestingly, in sound exposure studies, it appears less common to replicate sound treatments as well (Hurlbert, 1984; Kroodsma et al., 2001). However, sounds of one sound class can also vary due to individual, and potentially unique, traits of the sound source. Additionally, a sound recording can be confounded by additional unwanted and unnoticed sound sources (Slabbekoorn, 2013). Hence, exposing all animal subjects to a single sound treatment as representative of a class of sounds is referred to as simple pseudoreplication (Kroodsma et al., 2001). By using sufficient sound treatments in independent trials, the obtained results on the responses of animals are more likely to be representative of the response to this sound class in general (Slabbekoorn & Bouton, 2008). A suitable, truly replicated sound treatment design would

\* Corresponding author.

E-mail address: [jeroen.hubert@wur.nl](mailto:jeroen.hubert@wur.nl) (J. Hubert).

be either to expose each subject to a unique sound treatment (so, used once) of a class or to use multiple sound treatments to represent one class, and test multiple animals per sound treatment.

When multiple animals are individually tested using one of the sound treatments of a sound class, it is important to recognize that the animals that were exposed to the same sound treatment, even when individually tested, are not true independent replicates (Hurlbert, 1984; Kroodsmas et al., 2001). This means that the results cannot simply be pooled in the analysis (Machlis et al., 1985). Such unjustified pooling is referred to as sacrificial pseudoreplication (Hurlbert, 1984). To compare two sound classes that each comprises multiple treatments, one should ideally use models that can handle data with hierarchical or nested structures, such as a nested ANOVA or mixed-effect model (Kroodsmas et al., 2001; Millar & Anderson, 2004; Wiley, 2003). Attributing part of the observed variation to the individual treatments may actually increase the statistical power (Millar & Anderson, 2004).

The exposure design of an experiment affects not only the statistical methods that should be used, but also the conclusions that can be drawn from the study. For instance, speakers may not be able to mimic the sound field generated by the original source completely due to low-frequency limitations or the speaker being a point source. Consequently, studies that use speakers should be careful in generalizing their results to the sound of the original sound source (e.g. boat noise). Additionally, one should be aware that the actual source (e.g. a boat) may affect animals not only through sound, but also through stimulation in other modalities. Lastly, studies that have only used a single sound treatment to represent a sound class cannot draw conclusions about the sound class in general, because of the previously mentioned potential confounders in that sound treatment (Slabbekoorn, 2013).

For the current paper, we examined previously published underwater sound exposure studies and identified the sound treatment designs, analyses and interpretations. Using these results, we discuss to what extent pseudoreplication still occurs in this field and what the consequences are. The occurrence of pseudoreplication in terrestrial exposure studies has been quantified before and this has decreased over time (Kroodsmas, 1991; Kroodsmas et al., 2001). Now that underwater exposure studies are proliferating, it may be useful to map pseudoreplication here too. This commentary should provide guidance for all future (underwater) sound exposure experiments and can hopefully contribute to reducing pseudoreplication. Despite the focus of the literature survey on underwater sound exposure studies, all principles apply to in-air studies on terrestrial organisms too.

## METHODS

We used Web of Science (see search query in the Appendix) to find papers on underwater sound exposure studies, published from 2019 to 2023, in English. We screened the resulting records ( $N = 495$ ) on title and abstract. Records were eliminated if they did not contain sound exposure experiments, if they focused on humans or plants, and if they used stimuli that were not recorded underwater (e.g. cars, aeroplanes) or artificial stimuli only. The remaining records ( $N = 95$ ) were assessed for eligibility based on the full text: one record was excluded during this round because it lacked information regarding composition of playback sound stimuli and sound sources. Additional records ( $N = 10$ ) were added at this stage, gathered by looking up referenced articles from the earlier identified records. A flow diagram of the systematic literature search is provided in Appendix Fig. A1.

All included papers ( $N = 104$ ) were scored in the subsequent step. All studies compared two or more sound classes: for example, ambient sound (as control) and boat sound, or one or more sound

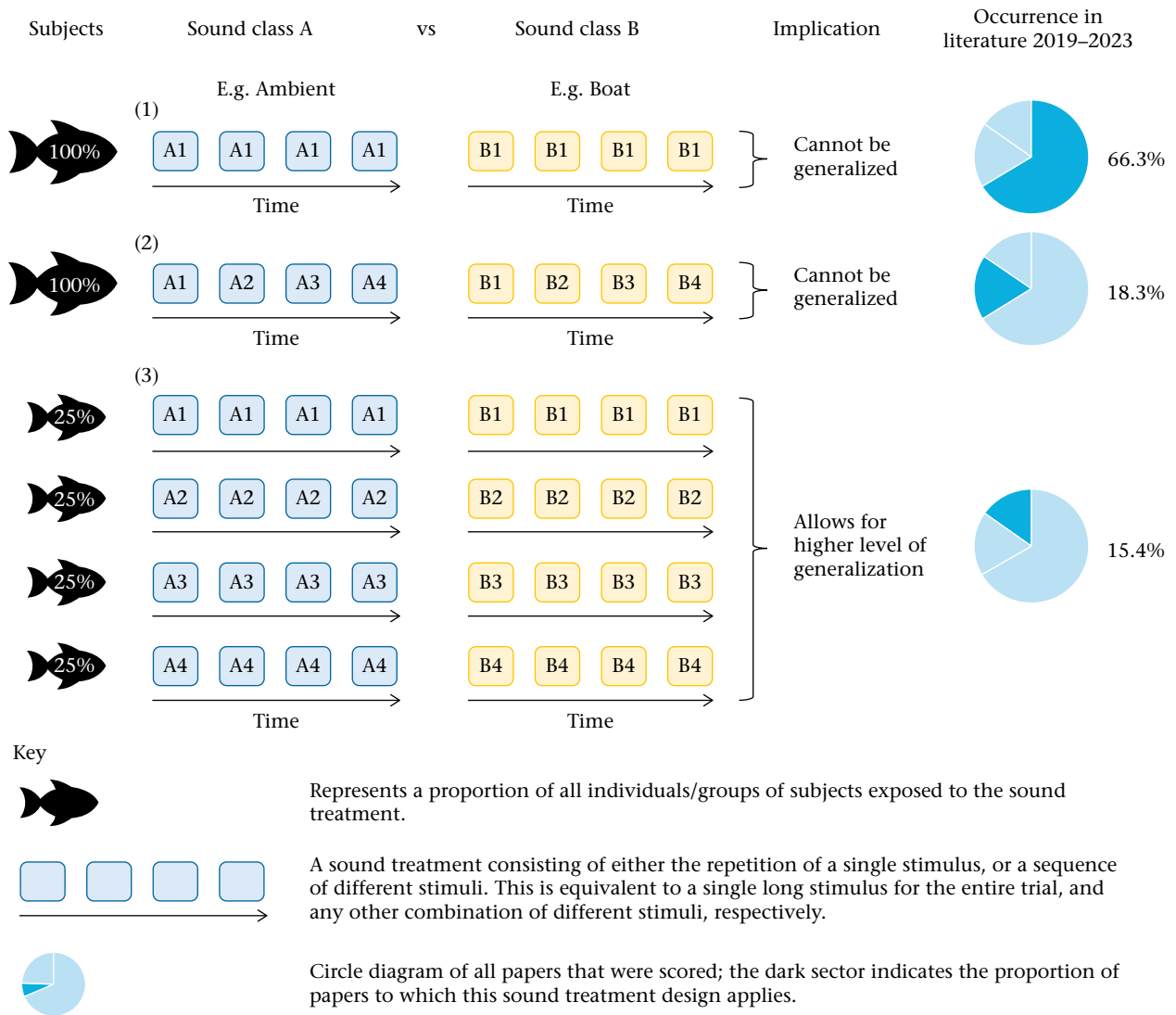
class(es) and silence. We scored whether only one sound treatment was used for each sound class (yes/no). In other words, were all subjects, per sound class, exposed to the same sound treatment? Sound classes that were just silence playback were ignored. We also scored how many unique stimuli (recordings or sources) were used to make the sound treatment(s) of each sound class. Multiple recordings of, or exposures to, one individual source were scored as one stimulus or treatment. Multiple treatments that used the same recordings (but e.g. in a different order) were regarded as one. When the number of stimuli was unknown (but clearly more than one per sound class) or uneven between sound classes, we scored the number as '>1'. Sound classes with artificial stimuli (e.g. white noise, pure tones) were ignored for the scoring. The two scores allowed us to differentiate between three different stimulus designs: (1) all subjects were exposed to the same sound treatment per sound class, and the sound treatment consisted of one stimulus only; (2) all subjects were exposed to the same sound treatment per sound class, and the sound treatment consisted of a combination of multiple stimuli; or (3) the subjects were exposed to different sound treatments per sound class, consisting of different stimuli. A number of papers consisted of multiple experiments, but in those cases, either only one experiment met the inclusion criteria or the different experiments fell into the same stimulus design category and was therefore treated as one study. For papers with the third treatment design, we scored whether replicated (reused?) sound treatments were taken into account in the statistical analyses. Additionally, we scored whether the subjects were exposed to sound using a speaker or by the original sound source. To get an impression of the generalizations made by the authors based on their studies, and whether these fitted their experimental designs, we checked the title and abstract of each paper and noted whether these contained any definitive conclusions about (1) the actual source of the sound, when the study had used playbacks, (2) only the sound of a source, when the study had used the source with potentially also stimuli in other modalities, and (3) the sound class as a whole. We present the results below, give examples and directly discuss the implications, informed through previous perspective papers.

## RESULTS AND DISCUSSION

### *Sound Treatment Design*

The systematic literature analysis of underwater sound exposure studies revealed that the majority of the studies were performed using a single sound treatment per sound class, each class consisting of one stimulus only (Fig. 1a). This means that the results of these studies are only valid for the sound treatments used and cannot be extrapolated beyond them. Depending on the goal of the study, and given that appropriate statistics are used, this may not be problematic. For example, if only one boat or boat type is allowed in a specific area, and the goal is to study the effects of this specific boat on the local ecosystem, then generalization beyond the tested treatments is not important. When such a study is performed using speaker playback of recordings, it is still recommended to use different recordings from the boat or preferably multiple boats of the specific type to dilute and control for potentially unknown biases in a recording. It is under debate whether a study examining a single stimulus should even use inferential statistics or rather descriptive statistics only (Cottenie & De Meester, 2003; Hurlbert, 1984; Oksanen, 2001). Most importantly, for both authors and readers of such papers, it should be clear that the results may only be valid for the stimuli used.

A study that highlighted the need for sufficient replicated sound treatments exposed foraging shore crabs, *Carcinus maenas*, to one of six ambient and six boat playbacks (Hubert et al., 2021). The time



**Figure 1.** Overview of three sound treatment designs to compare effects of two sound classes, the summarized implication, and the occurrence of such a design in the literature ( $N = 104$  included papers). The majority of the papers used a single sound treatment (per sound class), each consisting of a single stimulus (1). These results cannot be generalized beyond the sound treatment used because it may not be representative of the sound class in general, and a single recording may be confounded by other unwanted sources. Less than one-fifth of the papers used a single sound treatment (per sound class), each consisting of multiple stimuli (2). These results may be explained by just one of the stimuli in the treatment, so the same risks as for design 1 applies here. Only the smallest category of papers used multiple sound treatments (per sound class), consisting of different stimuli (3). The latter generally leads to results with the highest external validity, because the treatments more likely represent the sound class as a whole and the impact of confounded treatments is diluted.

the crabs used to reach the food was scored. This foraging time was similar during all six ambient playbacks. For five of six boat playbacks, the foraging time was faster than during the ambient playbacks, while one boat playback resulted in slower foraging times. If only the deviating boat playback had been used, the outcome of the study would have been completely reversed. The variability in responses to the boat playbacks used may also reflect the variability in response levels in situ with actual sound sources.

Almost a fifth of the papers used multiple stimuli per sound class but combined them in single sound treatments that were used for all subjects (Fig. 1b). It may appear as if stimuli were replicated; however, the response of the animals may be a result of just one of the stimuli. This means that the sound class is not truly replicated and that the results still cannot be extrapolated beyond the specific sound treatment. Again, it depends on the goal of the study, the statistical analysis and interpretation of the study whether this is really problematic, but the potential for generalization is probably

limited. Both designs (Fig. 1a and b) can be labelled as simple pseudoreplication (Kroodsma et al., 2001). It should be mentioned that some (field) studies are logistically very challenging, time-consuming and expensive (e.g. McQueen et al., 2022), especially in long-term studies. Despite the absence of replication for the sound treatment, these studies can still contribute significant value to the field due to the scarcity of such data, if one is aware of the limitations.

Only 16 of the 104 papers used multiple sound treatments per sound class, consisting of different stimuli (Fig. 1c). This is a design with true replication and more readily allows for generalization and extrapolation of the results, especially when there is little variation between the effects of individual sound treatments of a sound class. In that case, it is likely that untested sound treatments that fall acoustically within the variation of the tested sound treatments will elicit a similar effect. When different sound treatments of one sound class seem to cause opposing effects or the effects vary markedly, it is more complex to generalize the results.

Examination of the acoustic characteristics may inspire new hypotheses and experiments about sound impact. Additionally, it would be valuable to have information on the frequency of occurrence of each stimulus in situ.

A point for consideration in the latter category is that three of the 16 studies only used two sound treatments per sound class, and five only used three. Although two sound treatments are much better than just one, it may still be a poor representation of a sound class. Besides the limited replication within studies, several studies also reused treatments from earlier studies, leading to an overall limited pool of sound treatments.

### Analyses

From the studies that used multiple sound treatments per sound class ( $N = 16$ ), the majority (68.8%;  $N = 11$ ; Fig. 2a) pooled the results of all treatments from a class and did not statistically attribute potential variation in results to the individual treatments. This is referred to as sacrificial pseudoreplication and inappropriate pooling (Kroodsmas et al., 2001), violates the assumptions of independence of many statistical tests, and can actually reduce the statistical power of a test (Millar & Anderson, 2004). One study did not perform inferential statistics, one did not pool the results at all and three prevented sacrificial pseudoreplication by using each treatment once: Varola et al. (2021), Kok et al. (2021), and Sal Moyano et al. (2023) all used sufficient sound treatments to use each of them for one animal subject only. Within our sample, we found no studies that used different sound treatments multiple times and used statistical models that handle data with hierarchical or nested structures, which is another way to prevent sacrificial pseudoreplication.

### Speakers

The majority of the studies examined used speakers (74.0%; Fig. 2b) to expose subjects to sound, rather than the original sound source (24.0%; Fig. 2b), and two studies used both (1.9%; Fig. 2b). Speakers allow for an easy and controllable exposure and for

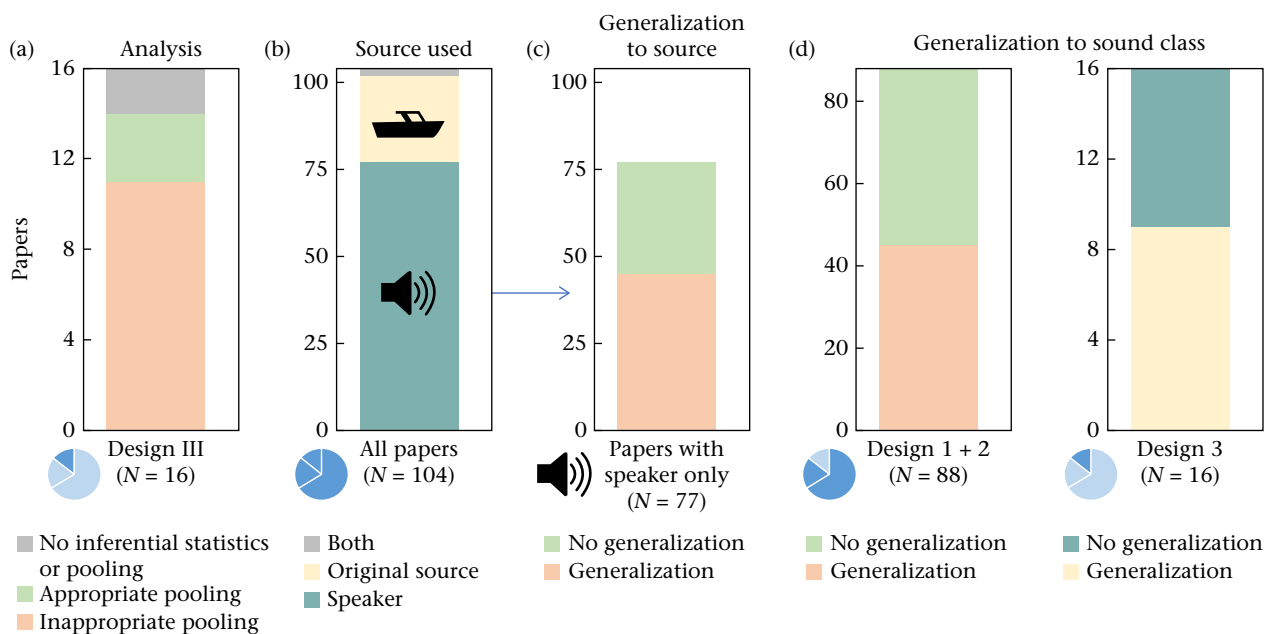
manipulation of sound treatments. On the other hand, it can be difficult to mimic original sound sources using a speaker. Underwater speakers are typically unable to produce low-frequency sound and are point sources whereas sources such as boats and pile driving are not. In laboratory settings, it is often not possible to use the original sound sources. In situ, it can be logistically challenging and expensive to replicate trials with the actual sound sources. Either way, one should be aware of the opportunities and limitations of speakers and actual sources which should match the goal of the experiment and interpretation of the results.

As an aside, two of the studies that used a speaker played back natural sounds and had the explicit goal to test the effect of the played back sound rather than the original sound in order to evaluate the efficacy of playbacks in reef restoration projects (McAfee et al., 2023; Williams et al., 2023). So, a speaker can also be the desired source.

### Interpretation

From the 77 papers that only used speakers for the exposure, 45 (58.4%; Fig. 2c) contained generalizations of the results to effects of the actual sound source in the title and/or abstract. This was, for instance, the case in papers that described experiments with playbacks of prerecorded boat noise, and derived conclusions about the effect of boat noise or anthropogenic noise in general from these. On the other hand, from the 25 papers that only used original sources, such as actual boats and pile driving events, five (20.0%) contained definitive conclusions about just the sounds produced by these sources in the title and/or abstract, even though the effect of other aspects of these sources could not be ruled out.

The title and/or abstract of 54 (53.5%) of the 104 papers contained definitive generalizations of results of the sound treatment(s) used to the sound class as a whole, for instance about boat noise in general. As mentioned before, studies using only a single sound treatment to represent a sound class should be cautious in drawing conclusions about the sound class in general, because of the treatment probably not being representative of the entire



**Figure 2.** (a) Type of pooling that was applied for papers with design 3 (see Fig. 1 for the three treatment designs). (b) Sound source used by all papers. The boat symbol is just an example of an original source. (c) Generalization to the original sound source by papers that used a speaker only. (d) Generalization of the sound treatment(s) used to the sound class. Note that most y-axes differ.



sound class (Slabbekoorn, 2013). However, of the 88 papers using only a single sound treatment per sound class, 45 (51.1%; Fig. 2d) contained definitive generalizations to the entire sound class in the title and/or abstract. Interestingly, an only slightly higher percentage (56.3%, nine of 16; Fig. 2d) of the papers that used multiple sound treatments per sound class contained definitive generalizations to the sound class as a whole in the title and/or abstract, even though, as mentioned earlier, this design more readily allows for extrapolation of the results to the entire sound class.

### Conclusion

Exposure studies with appropriate design and testing of sound stimuli can provide insight into sound impact, mitigation and animal cognition. To be able to extrapolate results from sound exposure studies beyond the specific sound treatments used in the study, multiple sound treatments representing a sound class must be used in separate sound exposure trials. The current systematic literature analysis shows that this is not yet common practice in underwater exposure studies, which limits the external validity of results. Avoiding pseudoreplication may be more challenging for long-term studies in the field and with original sound sources rather than speakers, all of which increase the ecological validity of the study; nevertheless, one should be aware of the risks and interpret the results accordingly. To enable researchers to replicate sound treatments, we call for generous sharing of (unused) recordings.

### Author Contributions

**Jeroen Hubert:** Writing – original draft, Visualization, Supervision, Methodology, Investigation, Conceptualization. **Max Willems:** Writing – review & editing, Visualization, Methodology, Investigation, Conceptualization. **Judith M. Varkevisser:** Writing – review & editing, Methodology, Investigation, Conceptualization.

### Data Availability

The data and R scripts are available as Supplementary Material.

### Declaration of Interest

The authors declare that they have no conflicts of interest to disclose regarding this research.

### Acknowledgments

J.H. was funded by an NWA grant (NWA.1236.18.004) from the Netherlands Organization of Scientific Research (NWO).

### Supplementary Material

Supplementary material associated with this article is available, in the online version, at <https://doi.org/10.1016/j.anbehav.2024.07.005>.

### References

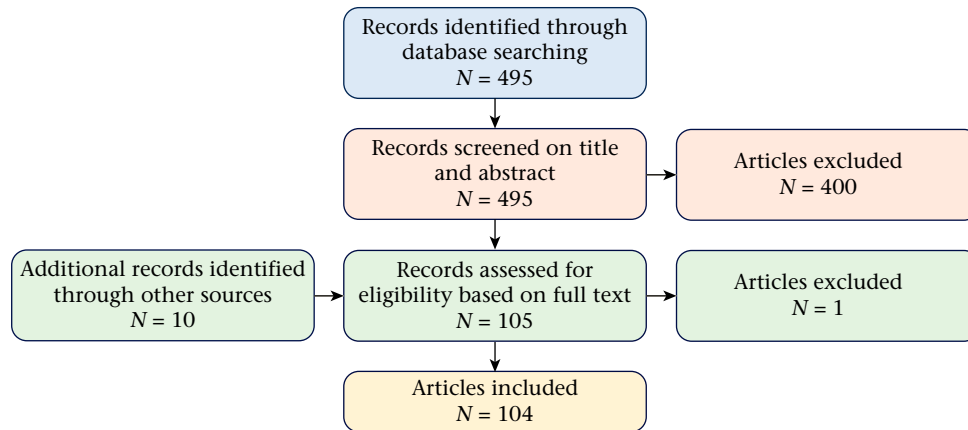
- Cottenie, K., & De Meester, L. (2003). Comment to Oksanen (2001): Reconciling Oksanen (2001) and Hurlbert (1984). *Oikos*, 100(2), 394–396. <https://doi.org/10.1034/j.1600-0706.2003.11953.x>
- Duarte, C. M., Chapuis, L., Collin, S. P., Costa, D. P., Devassy, R. P., Eguiluz, V. M., Erbe, C., Gordon, T. A. C., Halpern, B. S., Harding, H. R., Havlik, M. N., Meekan, M., Merchant, N. D., Miksis-Olds, J. L., Parsons, M., Predragovic, M., Radford, A. N.,

- Radford, C. A., Simpson, S. D., ... Juanes, F. (2021). The soundscape of the Anthropocene ocean. *Science*, 371(6529). <https://doi.org/10.1126/science.aba4658>
- EEA. (2020). *Environmental noise in Europe—2020*. <https://www.eea.europa.eu/publications/environmental-noise-in-europe>.
- Hubert, J., van Bemmelen, J. J., & Slabbekoorn, H. (2021). No negative effects of boat sound playbacks on olfactory-mediated food finding behaviour of shore crabs in a T-maze. *Environmental Pollution*, 270, Article 116184. <https://doi.org/10.1016/j.envpol.2020.116184>
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54(2), 187–211.
- Kok, A. C. M., van Hulsten, D., Timmerman, K. H., Lankhorst, J., Visser, F., & Slabbekoorn, H. (2021). Interacting effects of short-term and long-term noise exposure on antipredator behaviour in sand gobies. *Animal Behaviour*, 172, 93–102. <https://doi.org/10.1016/j.anbehav.2020.12.001>
- Kroodsma, D. E. (1991). How the mismatch between the experimental design and the intended hypothesis limits confidence in knowledge, as illustrated by an example from bird-song dialects. In M. Bekoff (Ed.), *Interpretation and explanation in the study of animal behavior* (pp. 226–245). Routledge.
- Kroodsma, D. E., Byers, B. E., Goodale, E., Johnson, S., & Liu, W.-C. (2001). Pseudoreplication in playback experiments, revisited a decade later. *Animal Behaviour*, 61(5), 1029–1033. <https://doi.org/10.1006/anbe.2000.1676>
- Kunc, H. P., & Schmidt, R. (2019). The effects of anthropogenic noise on animals: A meta-analysis. *Biology Letters*, 15(11). <https://doi.org/10.1098/rsbl.2019.0649>
- Machlis, L., Dodd, P. W. D., & Fentress, J. C. (1985). The pooling fallacy: Problems arising when individuals contribute more than one observation to the data set. *Zeitschrift für Tierpsychologie*, 68(3), 201–214. <https://doi.org/10.1111/j.1439-0310.1985.tb00124.x>
- MacKinlay, R. D., & Shaw, R. C. (2023). A systematic review of animal personality in conservation science. *Conservation Biology*, 37(1), Article e13935. <https://doi.org/10.1111/cobi.13935>
- McAfee, D., Williams, B. R., McLeod, L., Reuter, A., Wheaton, Z., & Connell, S. D. (2023). Soundscape enrichment enhances recruitment and habitat building on new oyster reef restorations. *Journal of Applied Ecology*, 60(1), 111–120. <https://doi.org/10.1111/1365-2664.14307>
- McQueen, K., Meager, J. J., Nyqvist, D., Skjæraasen, J. E., Olsen, E. M., Karlsen, Ø., Kvaldsheim, P. H., Handegard, N. O., Forland, T. N., & Sivle, L. D. (2022). Spawning Atlantic cod (*Gadus morhua* L.) exposed to noise from seismic airguns do not abandon their spawning site. *ICES Journal of Marine Science*, 79(10), 2697–2708. <https://doi.org/10.1093/icesjms/fsac203>
- Millar, R. B., & Anderson, M. J. (2004). Remedies for pseudoreplication. *Fisheries Research*, 70(2–3), 397–407. <https://doi.org/10.1016/j.fishres.2004.08.016>
- Oksanen, L. (2001). Logic of experiments in ecology: Is pseudoreplication a pseudoissue? *Oikos*, 94(1), 27–38. <https://doi.org/10.1034/j.1600-0706.2001.11311.x>
- Sal Moyano, M. P., Ceraulo, M., Luppi, T., Gavio, M. A., & Buscaino, G. (2023). Anthropogenic and biological sound effects on the maternal care behavior of a key crab species. *Frontiers in Marine Science*, 10, Article 1050148. <https://doi.org/10.3389/fmars.2023.1050148>
- Slabbekoorn, H. (2013). Songs of the city: Noise-dependent spectral plasticity in the acoustic phenotype of urban birds. *Animal Behaviour*, 85(5), 1089–1099. <https://doi.org/10.1016/j.anbehav.2013.01.021>
- Slabbekoorn, H., & Bouton, N. (2008). Soundscape orientation: A new field in need of sound investigation. *Animal Behaviour*, 76(4), e5–e8. <https://doi.org/10.1016/j.anbehav.2008.06.010>
- Tipton, E., Hallberg, K., Hedges, L. V., & Chan, W. (2017). Implications of small samples for generalization: Adjustments and rules of thumb. *Evaluation Review*, 41(5), 472–505. <https://doi.org/10.1177/0193841X16655665>
- Varola, M., Verga, L., Sroka, M. G. U., Villanueva, S., Charrier, I., & Ravignani, A. (2021). Can harbor seals (*Phoca vitulina*) discriminate familiar conspecific calls after long periods of separation? *PeerJ*, 9, Article e12431. <https://doi.org/10.7717/peerj.12431>
- WHO. (2011). Burden of disease from environmental noise—quantification of healthy life years lost in Europe. <https://www.who.int/publications/i/item/9789289002295>.
- Wiley, R. H. (2003). Is there an ideal behavioural experiment? *Animal Behaviour*, 66(3), 585–588. <https://doi.org/10.1006/anbe.2003.2231>
- Williams, B. R., McAfee, D., & Connell, S. D. (2023). Combining ecology and technology to kick-start oyster reef restoration. *Restoration Ecology*, 31(8), Article e13975. <https://doi.org/10.1111/rec.13975>

### Appendix

The search query used in Web of Science was as follows:

(TS=(playback) OR TS=(“sound exposure”) OR TS=(“noise exposure”) OR TS=(“anthropogenic noise”) OR TS=(“noise pollution”)) AND (TS=(marine) OR TS=(freshwater) OR TS=(underwater)) and Article (Document Types) and 2023 (Publication Years) OR 2022 (Publication Years) OR 2021 (Publication Years) OR 2020 (Publication Years) OR 2019 (Publication Years)



**Figure A1.** Flow diagram of the systematic literature search.