BMC Ecology and Evolution

RESEARCH Open Access

A harbor seal (*Phoca vitulina*) shows extensive respiratory control in sound production

Diandra Duengen[1,2*], Yannick Jadoul[1,4] and Andrea Ravignani[1,3,4,5*]

Abstract

The duration of animal vocalizations varies between and within species. Which mammals can learn to control this duration? Such respiratory production learning is a scarcely studied subcomponent of vocal learning. Here, we test the hypothesis that harbor seals (*Phoca vitulina*) are capable of respiratory production learning by testing whether a harbor seal can be trained to i) actively control its vocalization's duration in two directions (short and long), and ii) exceed the pre-experimental vocalization's duration (min = 0.202 s, max = 2.621 s). The seal learned to produce uninterrupted vocalizations spanning more than two orders of magnitude in duration, from 79 ms to 9.23 s. Our findings demonstrate a remarkable level of respiratory control in a harbor seal: this respiratory production learning encompasses an extensive range of sound durations and arises at a young age. Producing durations that span such a magnitude is hardly reported in the non-human animal literature; this capacity may be orthogonal to other vocal learning modules and should be tested in more species, both vocal production learners and non-learners.

Keywords Bioacoustics, Respiratory production learning, Breathing control, Vocal learning, Animal training

Background

Some species can learn to produce new vocalizations, while others exhibit a set of innate vocalizations that minimally change over time. Vocal production learning is the ability to learn new vocalizations or modify existing ones based on experience [34]. A typical example of such vocal learning species are songbirds and parrots [6, 48]. Mammalian examples include certain species of bats [39, 41], cetaceans [13, 20], elephants [60], and seals [14, 51, 58, 59]. Crucially, most vocal learning literature focuses on the acquisition of novel call types [13] or the convergence of vocalizations towards sound models [20]; instead, vocal learning through non-imitative mechanisms [10] or the modification of single vocal parameters [58] remain largely unexplored.

The production of most mammalian sounds involves three components: respiration, phonation, and filter [19, 27, 62]. Such sound production emerges through respiration, where air serves as a "medium" and both the pulmonary apparatus (e.g., lungs, respiratory muscles, diaphragm) and passive recoil forces generate subglottal pressure– the air pressure below the vocal folds necessary for sound production [23, 29]. As the air passes through the larynx, the laryngeal tissue begins to vibrate, which introduces flow variation into the expiratory airstream. The resulting pressure disturbances generate a sound with different degrees of periodicity [44], but typically containing a fundamental frequency (f_0; phonation)

*Correspondence:
Diandra Duengen
diandra.duengen@gmail.com
Andrea Ravignani
andrea.ravignani@uniroma1.it
[1]Comparative Bioacoustics Research Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Zoo Cleves ("Tiergarten Kleve"), Kleve, Germany
[3]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, The Royal Academy of Music, Aarhus, Denmark
[4]Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy
[5]Research Center of Neuroscience "CRiN-Daniel Bovet", Sapienza University of Rome, Rome, Italy

and harmonics. The airflow is then filtered through the supra-laryngeal vocal tract (filter), which generates formant frequencies [22].

The respiratory, phonatory, and filter components entail a complex mapping of interacting mechanisms [24, 26, 42, 47, 50, 53]. However, while phonation and filter typically shape the spectral properties of a sound, respiration and glottal resistance mainly influence its duration and amplitude [33]. Many mammals, including humans, modify single features, such as f_0 [41], formants [25, 58] or duration [43], which can encode information when communicating.

As for the general case of sound production, vocal production learning spans three levels: respiratory, phonatory, and filter production learning, all involving a different level of control over their corresponding apparatus [34, 49]. Respiratory production learning is suggested to have preceded filter and phonatory production learning [34], and temporal vocal parameters are easier to adjust than spectral parameters [21, 34]. Respiratory production learning requires basic control of the respiratory system (and vocal fold adduction for the case of voiced sounds); altering frequency parameters demands close coordination between the respiratory and phonatory systems [34]. For example, human speech production requires solid respiratory control [43].

Respiratory production learning occurs in species that are vocal production learners and non-learners [46, 61]; hence respiratory production learning could be found in more species than those showing filter or phonatory production learning. Consequently, investigating such learning provides a comparative insight into vocal learning. Durational variation can emerge as a by-product of other vocal adaptations. When trained to vocalize above a certain frequency threshold, lesser spear-nosed bats (*Phyllostomus discolor*) modified not only the spectral parameters of their calls but also the calls' duration and amplitude [40]. Conversely, in a similar experiment, harbor seals (*Phoca vitulina*) did not modify the duration of their calls [63]. Indirect evidence of respiratory production learning from durational adjustments may be confounded by emphasis on other acoustic parameters. So, how common is respiratory production learning across species?

Harbor seals are an ideal species to address this and related questions. They show some vocal learning capabilities [14, 17, 51, 55], with some control over the filter and phonatory level during sound production [25, 51, 63]. However, control over the third component– breathing– remains unexplored. This gap is intriguing, as breathing control in sound production may build upon seals' physiological diving adaptations [4, 12]. As part of their adaptation to diving, harbor seals have excellent respiratory control. From a very young age, they can hold their breath while diving for extended periods [36, 37]. This breathing control facilitates closing off the airways [1], i.e., controlling air leaving or entering the lungs. Phocid seals typically exhale before diving which reduces buoyancy [54]. Interestingly, phocid lungs are not very big but have comparably large oxygen storage capacities [8, 9]. Despite their remarkable diving and breathing physiology, it is unknown whether harbor seals exert their fine-tuned respiratory control during sound production.

Does the harbor seal, a species with elaborate breathing control and possible phonatory and filter production learning, show respiratory production learning? Here, we test if a harbor seal can control the duration of its vocalizations in two separate directions (short and long), thereby probing respiratory production learning.

Material & methods

Aim of the study

To test respiratory production learning, we trained a juvenile male harbor seal to produce short and long vocalizations when presented with a discriminative stimulus (S_D). We first trained the association between two arm gestures and short vs. long vocalizations. We then shaped the vocalizations' durations in a series of training sessions. Finally, the seal's ability for respiratory production learning was tested in a controlled experiment.

Experimental setup

A one-and-a-half-year-old male harbor seal participated in the study. This young male was housed with four conspecifics in a 230,000-L freshwater tank and 300 m^2 research enclosure at Zoo Cleves, Germany. We chose this animal to test respiratory production learning, while two other conspecifics, which also produced in-air vocalizations [15, 17], were tested in different tasks. All experiments were conducted on land in a research enclosure, neighboring the main enclosure. This research enclosure featured a 3 × 5 × 2.3 m carport above a 13.2 m^2 paved platform (see Fig. 1). Audio recordings were collected with a Zoom H6 digital recorder connected to a Sennheiser ME-67 unidirectional microphone (frequency response: 40–20.000 Hz ± 2.5 dB, sampling rate: 48 kHz) covered by a foam windshield. The microphone pointed straight towards the seal at one meter.

Training and testing procedure

We trained the seal through operant conditioning, reinforcing correct behavior positively with pre-cut pieces of fish. Incorrect behavior resulted in a three-second time-out. Initially, we trained the association of different movements of the right arm (discriminative stimulus, S_D) with different durations. Concretely, a leftward movement (i.e., towards the body) was trained to elicit a short vocalization, while a rightward movement (i.e., away

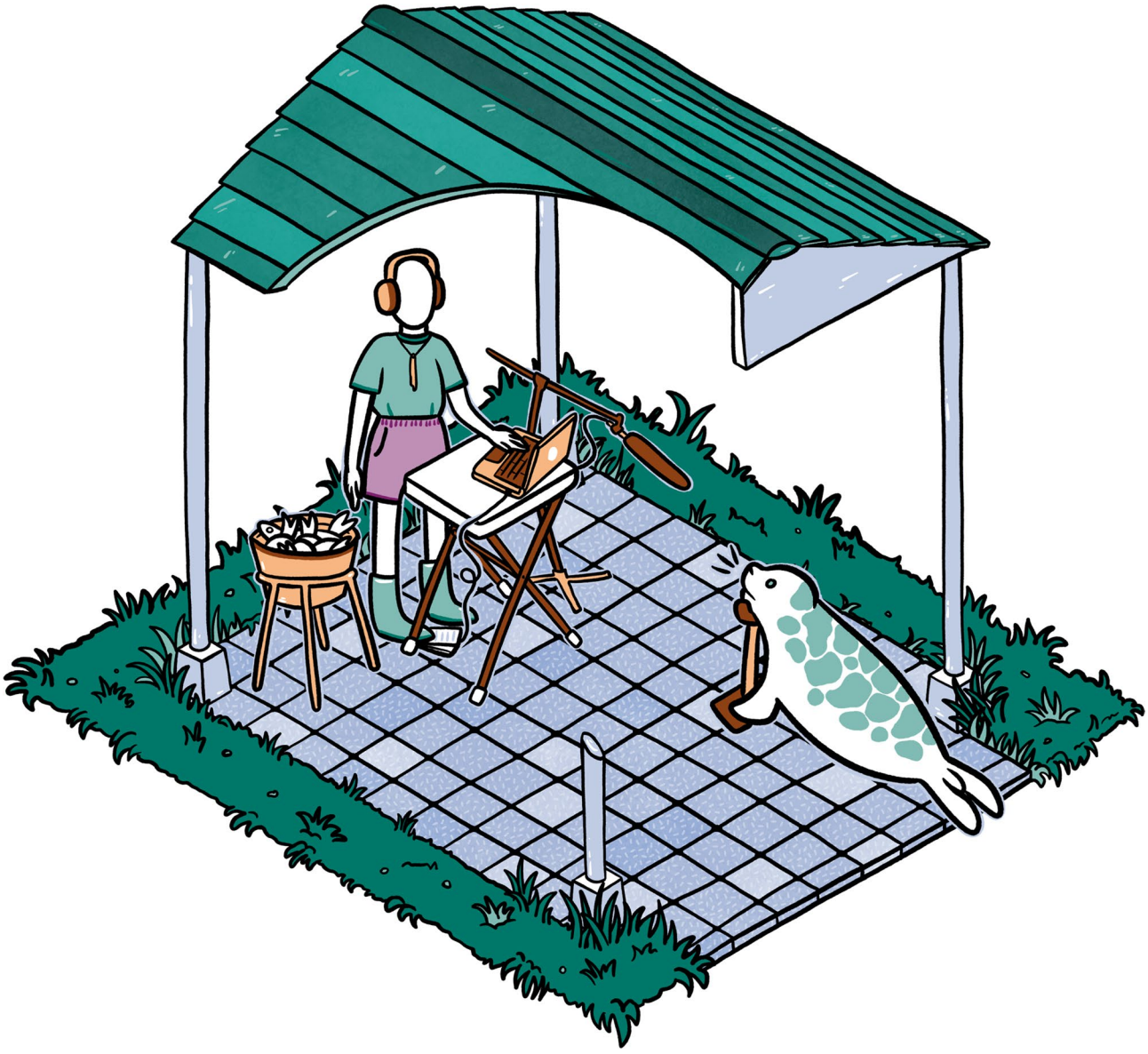

Fig. 1 The seal is positioned on its station in front of the microphone and recorder, which records the seal's vocal output. The experimenter stands in front of the laptop on which our custom-made application runs to assess the recorded response (see Fig. S1) of the seal

from the body) elicited a long vocalization. The experiment was conducted between September 2022 and January 2023– outside of the breeding season of harbor seals– and featured three phases.

First, the vocalization's durations were *subjectively* assessed by the trainer. If the trainer presented the S_D for "short", the seal was exclusively rewarded for a vocalization perceived as shorter than usual. Analogously, when the trainer presented the S_D for "long" the seal was exclusively rewarded for a longer vocalization. The seal participated in 21 sessions (on average 26.5 trials/session).

Next, to more *objectively* shape shorter and longer vocalizations, the vocalization's duration was directly displayed to the experimenter via a custom-made, Python-based computer application (integrating acoustic analyses through Parselmouth; [5, 31, 32]). For this, the experimenter pushed a foot pedal to start real-time sound analysis and released the pedal to stop it. The seal participated in 14 sessions (average 29.5 trials/session). Once the seal responded to each S_D with a vocalization, testing started.

The third phase consisted of the actual experiment, where we tested the seal's ability to adjust the duration of its vocalizations in alternating sessions, i.e., each experimental session consisted of only short or long trials, referred to below as $\text{sessions}_{\text{short}}$ and $\text{sessions}_{\text{long}}$, respectively. Initially, the seal was asked to produce shorter and longer vocalizations than a pre-determined

duration threshold of 0.611 s. Crucially, the seal had to respond with short or long vocalizations upon presentation of the S_D for "short" and "long". In other words, any vocalization below 0.611 s was correct in a session$_{short}$ and any vocalization above 0.611 s was correct in a session$_{long}$. This initial threshold of 0.611 s equaled the seal's pre-experimental median duration (based on a subsample of every 7th pre-experimental vocalization). A response was counted as correct if its measured duration surpassed a session's threshold (i.e., was shorter or longer than the threshold in a respectively short or long session). The threshold was adjusted once the seal reached the learning criterion: 80% correct choices in two consecutive sessions. The new threshold was then determined by calculating the median value of all responses in those last two sessions. For example, once the seal exceeded the duration of 0.611 s in at least 80% of all responses in two consecutive sessions$_{long}$, the median duration of the responses in the successful sessions (e.g., 0.768 s) was calculated and served as a new threshold. In the subsequent sessions$_{long}$ the seal was only rewarded for vocalizations longer than this new threshold until the learning criterion was reached again. The seal had a total of 110 sessions, each with 20 valid trials.

All responses were semi-automatically assessed through our application (Fig. S2). To start a trial, the experimenter pushed a foot pedal, presented the S_D, and released it once the seal had finished vocalizing. This indicated a rough time window in the recording around the seal's vocalization; based on this estimate, the application detected the vocalization, measured its duration, and compared it to the session's threshold (see "Automatic Detection of Vocal Responses" in the Supplement). As an additional fail-safe mechanism, the application presented the spectrogram of the recording and its corresponding measurement to the experimenter. If the measured duration surpassed the threshold, the assessment was marked as correct; otherwise, the trial was marked as incorrect. Background noise or other external disruptions, causing a false measurement, triggered an invalid trial which was repeated right away.

Acoustic analysis

The custom software analyzed the recorded vocalizations in real time and displayed an annotated spectrogram (Fig. S2). During the training phase, the application presented the spectrogram and measured the duration to help the experimenter decide which vocalizations to reinforce. During testing, the application additionally compared the measurement of each trial's vocalization to the session's pre-set threshold and proposed whether the trial was correct or incorrect. The experimenter then had three options: i) accept the application's decision, ii) reject the application's decision, or iii) repeat the trial. After the experiment, all durations were manually extracted to compare them to the application's automatic measurements.

Pre-experimental vocal repertoire

The pre-experimental repertoire comprises all vocalizations recorded before this experiment started (September 2022– November 2022), encompassing two vocalization types. These were pre-trained as part of the daily training routine. Concretely, the seal was trained to produce any type of sound and was directly rewarded whenever it did. This led to the emergence of vocal behavior under stimulus control (for more information on this type of procedure, see [17]). For this study, one of the two trained vocalizations was selected and trained to be produced upon presentation of the S_D (Fig. S1). The vocalization's pre-experimental duration ranged from 0.202 s to 2.621 s (median: 0.611 s, Tab. S1).

Statistical analyses

The experiment resulted in three different datasets: one dataset before the experiment– the pre-experimental vocalizations' durations– and two datasets of the experimental data– the automatically measured durations of vocal responses and manually measured durations, excluding occasional trials with more than one recorded response. To test whether the seal successfully changed the duration of its vocalizations, we statistically compared the pre-experimental vocalizations to the manually annotated experimental vocalizations using a Mann–Whitney U test. Similarly, we statistically tested the change in duration between sessions with different thresholds. We tested the validity of the automatically measured durations by correlating them (Spearman's rank correlation coefficient) with the manually annotated vocal responses (see Fig. S3).

Results

Modification of duration throughout the experiment

Before the experiment, the duration of the seal's vocalizations was 0.652 s ± 0.243 s (see Tab. S1). Over time, the seal learned to adjust its vocalizations to have an increasingly shorter or longer duration (Fig. 2). Durations in the final session$_{long}$ (average: 5.336 s ± 1.312 s) significantly differed from the pre-experimental durations (Mann–Whitney U test, $U = 0.0$, $p < 0.0001$). Durations in the final session$_{short}$ (average: 0.160 s ± 0.042 s) also significantly differed from the pre-experimental durations (Mann–Whitney U test, $U = 15{,}443.0$, $p < 0.0001$). Grouping the responses by their session's threshold reveals significant changes in duration between every pair of consecutive thresholds (except between session$_{short}$ thresholds 0.611 s and 0.272 s). Even the response durations measured at the start of the testing phase (session$_{short}$ and session$_{long}$

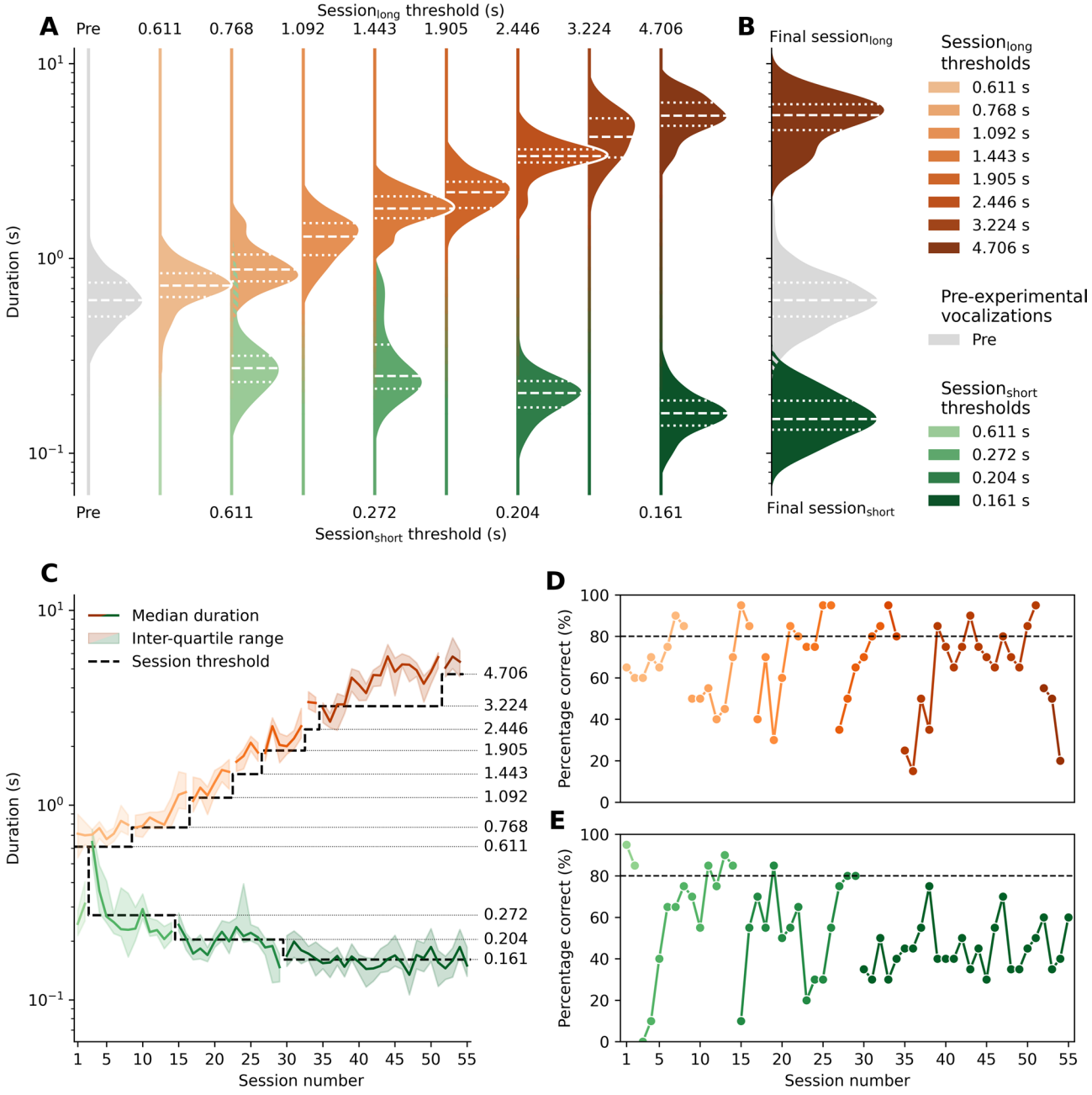

Fig. 2 The experimental results show that the seal successfully learned to gradually increase and decrease the duration of its vocalizations. **A**) Grouped by session threshold, the sessions$_{long}$ durations (orange) increase between successive thresholds, and vice versa for sessions$_{short}$ (green). These distributions become increasingly distinct from the pre-experimental vocalizations' durations (grey). The dashed lines indicate a distribution's median, the dotted lines the first and third quartiles. Note the logarithmic scaling of the y-axis, spanning over two orders of magnitude. **B**) At the end of the experiment, there is virtually no overlap between the final session$_{short}$ and session$_{long}$ distributions and the pre-experimental one. **C**) Similarly, the median and inter-quartile range of the responses per session display the diverging vocal durations. **D**) The learning curve of the sessions$_{long}$ shows the seal's respiratory production learning progress over time. **E**) The learning curve of the sessions$_{short}$ demonstrates the seemingly higher difficulty in consistently producing shorter vocalizations as the threshold lowers, especially for the final threshold of 0.161 s

threshold 0.611 s) already significantly differ from the pre-experimental durations (see Fig. 2A and Table S1).

Learning curves

The seal learned to respond to two different S_Ds with the same vocalization of different durations (short and long): when shown the S_D for "short", the seal emitted a short vocalization, and when shown the S_D for "long", a long vocalization. The seal learned to gradually decrease (short) or increase (long) the duration of its vocalization. This was achieved in 9.7 ± 5.6 $sessions_{short}$ and 7.3 ± 4.4 $sessions_{long}$ per threshold (Fig. 2D).

Discussion

The harbor seal successfully learned to increase and decrease the duration of a selected vocalization. By the end of the experiment, its shortest and longest produced vocalizations differed by over two orders of magnitude, demonstrating an exceptional level of breathing control during vocal production (see below for a discussion on the role of glottal adduction). The pre-experimental durational range (minimum: 0.20 s; maximum: 2.62 s) was extended to a minimum and maximum duration of 0.079 s and 9.23 s, respectively. These results make a clear case of respiratory production learning in a harbor seal.

Overall, the seal reached each intermediate learning criterion (two sessions with ≥ 80% correct responses) rather quickly. On average, the seal took 9.7 ± 5.6 $sessions_{short}$ and 7.3 ± 4.4 $sessions_{long}$. The last two thresholds (0.161 s for "short", and 4.706 s for "long") were not learned before the experiment ended. Perhaps, the observed 0.079 s are the minimum duration the seal can reach. We noted that the vocalizations dropped in intensity as they got shorter. These short durations may be difficult to detect acoustically with our experimental setup, which may explain the comparably high number (n = 26) of $sessions_{short}$. We suggest investigating such matters in an acoustic chamber to circumvent such limitations. While the last $session_{short}$ threshold (0.161 s) took longer to reach than usual, we don't see any indication of that within the final $session_{long}$ results (4.706 s). On the contrary, with only four sessions completed, the seal produced some very long vocalizations. With more sessions, the seal would have likely produced even longer vocalizations. This is supported by the many values exceeding the last threshold, most notably a maximum duration of 9.228 s and the average number of $sessions_{long}$ (7.6 ± 4.5, see above). However, since the experiment ended earlier– for reasons out of our control– the seal may not have had the opportunity to reach its maximally possible vocal duration.

Some primates, carnivores and bats show respiratory control [40, 46, 49, 61]. Our experiment finds a drastically larger change in durational range than observed, for instance, in lesser spear-nosed bats [40]. In that study, one bat showed an altered median duration range of 0.054 s, compared to the pre-experimental range of 0.04 s. The experiment, however, did not aim to shorten or lengthen the bats' vocalizations; such changes occurred as a by-product of another task, and likely underestimate the actual respiratory control in these bats. In contrast, Sutton et al. [61] directly tested whether (*Macaca mulatta*) could increase call duration and found that they did so up to three times the pre-experimental value (inferred from Fig. 3 in [61]). The harbor seal reported here demonstrated a durational increase of almost nine times on average, and a decrease of almost four, within just five weeks. Compared to previous work, we show specific control over the onset and offset of sound duration. Crucially, the seal learned to push the extent of one vocalization into two separate directions. It produced a range of vocal durations spanning over two orders of magnitudes. This range strongly exceeds the mammalian examples above [40, 61]. Overall, this indicates exceptional breathing control of in air sound production in a harbor seal.

Our results are relevant in the context of the evolution of speech: vocal learning is a crucial building block of speech. Respiratory production learning has been proposed as an evolutionary early form of vocal production learning and an instance of usage learning, as it primarily involves controlling the onset and offset of vocalizations [35], which might place this ability as an intermediate or transitional form between contextual learning and production learning [7]. The respiratory production learning we find in a mammal with a vocal apparatus remarkably similar to ours [52] may point to shared advanced respiratory control, perhaps convergently evolved in humans. Such control is essential to produce diverse speech sounds in humans [56] and to manage the timing of calls in harbor seal pups [2]. Our findings add additional evidence that harbor seals possess this capacity and can exert an extensive level of control over their vocalizations' duration.

Our results also demonstrate that a harbor seal can learn to produce vocalizations of longer and shorter durations through a non-imitative experimental approach, providing an alternative to the classical vocal imitation studies [58, 59]. Past studies found vocal flexibility and indications of vocal production learning in harbor seals [51, 63]. Given the lack of conclusive evidence, this species' ability and extent of vocal production learning is still debated [11]. Here, we have shown that a harbor seal can be conditioned to show novel temporal vocal features, thereby demonstrating respiratory production learning.

Our findings make it evident that the harbor seal was able to control the duration of its vocalizations. However, the specific mechanisms underlying sound production

in this case remain unclear. Typically, the duration of a vocalization does not depend on the pulmonary apparatus alone but also involves the larynx [47, 57]. There, the duration of phonation (i.e., the production of a voiced sound) depends mainly on the phonatory lung volume (the amount of air that is available to produce a voiced sound) and the air flow rate (the moved air volume over time; [3]). The phonatory lung volume is restricted to vital capacity, i.e., the maximum amount of air that can be exhaled after having inhaled a maximum amount of air. The airflow rate is influenced through subglottal pressure, vocal fold adduction, and the resulting glottal adduction [27]. Which roles these physiological parameters played in our sound production experiment are unfortunately unclear. The seal vocalized with its nostrils open and mouth closed; the resulting sounds appeared to be voiceless (see Fig. S1). This suggests that the sounds were generated through nasal airflow rather than laryngeal phonation. Due to a lack of direct measurements (e.g., spirometry, airflow analysis), we cannot definitively determine the roles of the nose, larynx, or supralaryngeal vocal tract in the production of these sounds. Future research should identify such roles, and include measurements of vital capacity and airflow rate [18], as well as directly investigate laryngeal involvement [28, 64] to clarify the contributions of different sound-producing structures in harbor seals. Future research should continue probing the limits of pinnipeds' vocal production learning, ideally on the phonatory and filter level. Like our approach, one could test whether stepwise increments of vocal parameter values can lead to a different post-experimental parameter distribution, providing clear indications of vocal learning on these levels. Future research should include a bigger sample size, to compare among age classes and sexes, as it remains unclear whether vocal learning abilities depend on age and/or sex [14, 16]. Besides measurements of pulmonary capacity and airflow rate, behavioral data should be linked to the underlying neurobiology in this and other pinniped species [12, 30, 38, 45].

Conclusion

We demonstrated, in a controlled experiment, a harbor seal's ability for respiratory production learning, a proposed evolutionary precursor to vocal production learning. The seal not only mastered control over diverse durations of its vocalizations but also learned to manipulate sound duration in two distinct directions (short and long). Remarkably, the seal achieved durations that surpassed the pre-experimental range and produced vocalizations that covered durations of two orders of magnitude, showcasing an impressive level of vocal control and adaptability.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12862-025-02404-9.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We thank Zoo Cleves for their readiness to cooperate and contribute to research. We are grateful to the harbor seal Robert and everyone who supported our work in this study, in particular Lara Schmitt and Martin Polotzek.

Authors' contributions

Conceptualization: DD, AR, YJ; Data Curation: DD; Formal Analysis: DD, YJ; Funding Acquisition: AR; Investigation: DD; Methodology: DD, AR, YJ; Project Administration: DD; Resources: DD, AR; Software: YJ, DD; Supervision: AR; Validation: DD, YJ; Visualization: DD, YJ; Writing– Original Draft Preparation: DD; Writing– Review & Editing: DD, AR, YJ.

Funding

Open Access funding enabled and organized by Projekt DEAL. Center for Music in the Brain was funded by the Danish National Research Foundation (DNRF117). The Comparative Bioacoustics Group was funded by Max Planck Group Leader funding to AR. YJ and AR are supported by the European Union (ERC, TOHR, 101041885) and by the HFSP research grant RGP0019/2022 (https://doi.org/https://doi.org/10.52044/HFSP.RGP00192022.pc.gr.153618).

Data availability

The research data used for this article has been uploaded in this submission.

Declarations

Ethics approval and consent to participate

All training and testing were conducted according to the requirements of the Landesamt für Natur, Umwelt und Verbraucherschutz (LANUV) NRW, Germany, section animal experiment affairs. Upon detailed discussions with the LANUV, the need for approval was waived (file number Az.81–04.78). Informed consent was obtained from all owners of the animals used in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 21 February 2025 / Accepted: 5 June 2025
Published online: 02 September 2025

References

1. Adams A, Vogl W, Dawson C, Raverty S, Haulena M, Skoretz SA. Laryngeal and soft palate valving in the harbour seal (*Phoca vitulina*). J Exp Biol. 2020;223(20):jeb230201. https://doi.org/10.1242/jeb.230201.
2. Anichini M, De Reus K, Hersh TA, Valente D, Salazar-Casals A, Berry C, Keller PE, Ravignani A. Measuring rhythms of vocal interactions: a proof of principle in harbour seal pups. Philos Trans R Soc B. 2023;378(1875):20210477. https://doi.org/10.1098/rstb.2021.0477.
3. Baken RJ. Clinical measurement of speech and voice. London: Taylor & Francis Ltd; 1987.
4. Blix AS. Adaptations to deep and prolonged diving in phocid seals. J Exp Biol. 2018;221(12):jeb182972. https://doi.org/10.1242/jeb.182972.
5. Boersma P, Weenink D. (2022). Praat: doing phonetics by computer In Glot International (Version 6.2.12) retrieved 17 April 2022 from http://www.praat.org/.
6. Bradbury JW, Balsby TJ. The functions of vocal learning in parrots. Behav Ecol Sociobiol. 2016;70(3):293–312. https://doi.org/10.1007/s00265-016-2068-4.

7. Brenowitz EA, Beecher MD. An ecological and neurobiological perspective on the evolution of vocal learning. Front Ecol Evol. 2023;11:1193903. https://doi.org/10.3389/fevo.2023.1193903.
8. Burns J, Costa D, Frost K, Harvey J. Development of body oxygen stores in harbor seals: effects of age, mass, and body composition. Physiol Biochem Zool. 2005;78(6):1057–68. https://doi.org/10.1086/432922.
9. Burns J, Lestyk K, Folkow L, Hammill M, Blix A. Size and distribution of oxygen stores in harp and hooded seals from birth to maturity. J Comp Physiol B. 2007;177:687–700. https://doi.org/10.1007/s00360-007-0167-2.
10. Carouso-Peck S, Goldstein MH. Female social feedback reveals non-imitative mechanisms of vocal learning in zebra finches. Curr Biol. 2019;29(4):631–636. e633. https://doi.org/10.1016/j.cub.2018.12.026.
11. Casey C, Sills JM, Knaub S, Sotolotto K, Reichmuth C. Lifelong Patterns of Sound Production in Two Seals. Aquatic Mammals. 2021;47(5):499–514. https://doi.org/10.1578/AM.47.5.2021.499.
12. Cook P, Rouse A, Sawyer E, Miller K, Berns G. Neurobiological adaptations supporting vocal plasticity have accumulated in the marine carnivores. 2025. https://doi.org/10.21203/rs.3.rs-6222519/v1.
13. Crance JL, Bowles AE, Garver A. Evidence for vocal learning in juvenile male killer whales, *Orcinus orca*, from an adventitious cross-socializing experiment. J Exp Biol. 2014;217(8):1229–37. https://doi.org/10.1242/jeb.094300.
14. Duengen D, Fitch WT, Ravignani A. Hoover the talking seal. Curr Biol. 2023;33(2):R50–2. https://doi.org/10.1016/j.cub.2022.12.023.
15. Duengen D, Jadoul Y, Ravignani A. Vocal usage learning and vocal comprehension learning in harbor seals. BMC Neurosci. 2024;25(1):48. https://doi.org/10.1186/s12868-024-00899-4.
16. Duengen D, Ravignani A. The paradox of learned song in a semi-solitary mammal. Ethology. 2023. https://doi.org/10.1111/eth.13385.
17. Duengen D, Ravignani A. Training Harbor Seals to Participate in Vocal Learning Experiments in a Zoo. bioRxiv.2024;2024.2008.2027.609954. https://doi.org/10.1101/2024.08.27.609954.
18. Fahlman A, Moore MJ, Garcia-Parraga D. Respiratory function and mechanics in pinnipeds and cetaceans. J Exp Biol. 2017;220(10):1761–73. https://doi.org/10.1242/jeb.126870.
19. Fant G. Acoustic theory of speech production. 2nd ed. The Hague: Mouton; 1970.
20. Favaro L, Neves S, Furlati S, Pessani D, Martin V, Janik VM. Evidence suggests vocal production learning in a cross-fostered Risso's dolphin (*Grampus griseus*). Anim Cogn. 2016;19(4):847–53. https://doi.org/10.1007/s10071-016-0961-x.
21. Fitch WT. The evolution of speech: a comparative review. Trends Cogn Sci. 2000;4(7):258–67. https://doi.org/10.1016/S1364-6613(00)01494-7.
22. Fitch WT, Anikin A, Pisanski K, Valente D, Reby D. Formant analysis of vertebrate vocalizations: achievements, pitfalls, and promises. BMC Biol. 2025;23(1):92. https://doi.org/10.1186/s12915-025-02188-w.
23. Fitch WTS. Production of Vocalizations in Mammals. In Encylopedia of Language and Linguistics. 2006; pp. 115–121. Elsevier (incl. Pergamon). https://doi.org/10.1016/B0-08-044854-2/00821-X.
24. Goller F. Respiratory contributions to birdsong—evolutionary considerations and open questions. Philos Trans R Soc B. 2025;380(1920):20230431. https://doi.org/10.1098/rstb.2023.0431.
25. Goncharova M, Jadoul Y, Reichmuth C, Fitch WT, Ravignani A. Vocal tract dynamics shape the formant structure of conditioned vocalizations in a harbor seal. Annals New York Acad Sci. 2024;1538(1):107–16. https://doi.org/10.1111/nyas.15189.
26. Hage SR. Breathing control of vocalization. Science. 2024;383(6687):1059–60.
27. Herbst CT. A review of singing voice subsystem interactions—toward an extended physiological model of "support." J Voice. 2017;31(2):249 e213-249. e219. https://doi.org/10.1016/j.jvoice.2016.07.019.
28. Herbst CT. Electroglottography–an update. J Voice. 2020;34(4):503–26. https://doi.org/10.1016/j.jvoice.2018.12.014.
29. Hixon TJ. Respiratory function in speech and song. Boston: College-Hill; 1987.
30. Hoeksema N, Verga L, Mengede J, Van Roessel C, Villanueva S, Salazar-Casals A, Rubio-Garcia A, Curcic-Blake B, Vernes SC, Ravignani A. Neuroanatomy of the grey seal brain: bringing pinnipeds into the neurobiological study of vocal learning. Philos Trans R Soc B. 2021;376(1836):20200252. https://doi.org/10.1098/rstb.2020.0252.
31. Jadoul Y, De Boer B, Ravignani A. Parselmouth for bioacoustics: automated acoustic analysis in Python. Bioacoustics. 2023;33:1–17. https://doi.org/10.1080/09524622.2023.2259327.
32. Jadoul Y, Thompson B, De Boer B. Introducing Parselmouth: A Python interface to Praat. J Phon. 2018;71:1–15. https://doi.org/10.1016/j.wocn.2018.07.001.
33. Janik VM, Knörnschild M. Vocal production learning in mammals revisited. Philos Trans R Soc B. 2021;376(1836):20200244. https://doi.org/10.1098/rstb.2020.0244.
34. Janik VM, Slater PJ. The different roles of social learning in vocal communication. Anim Behav. 2000;60(1):1–11. https://doi.org/10.1006/anbe.2000.1410.
35. Janik V, Slater PJB. Vocal learning in mammals. Adv Study Behav. 1997;26:59–99.
36. Jørgensen C, Lydersen C, Brix O, Kovacs KM. Diving development in nursing harbour seal pups. J Exp Biol. 2001;204(22):3993–4004. https://doi.org/10.1242/jeb.204.22.3993.
37. Kaczmarek J, Reichmuth C, McDonald BI, Kristensen JH, Larson J, Johansson F, Sullivan JL, Madsen PT. Drivers of the dive response in pinnipeds; apnea, submergence or temperature? J Exp Biol. 2018;221(13):jeb176545. https://doi.org/10.1242/jeb.176545.
38. Kendall-Bar JM, Williams TM, Mukherji R, Lozano DA, Pitman JK, Holser RR, Keates T, Beltran RS, Robinson PW, Crocker DE. Brain activity of diving seals reveals short sleep cycles at depth. Science. 2023;380(6642):260–5. https://doi.org/10.1126/science.adf0566.
39. Knörnschild M. Vocal production learning in bats. Curr Opin Neurobiol. 2014;28:80–5. https://doi.org/10.1016/j.conb.2014.06.014.
40. Lattenkamp EZ, Vernes SC, Wiegrebe L. Volitional control of social vocalisations and vocal usage learning in bats. J Exp Biol. 2018;221(Pt 14):jeb180729. https://doi.org/10.1242/jeb.180729.
41. Lattenkamp EZ, Vernes SC, Wiegrebe L. Vocal production learning in the pale spear-nosed bat*Phyllostomus discolor*. Biol Lett. 2020;16(4):20190928. https://doi.org/10.1098/rsbl.2019.0928.
42. MacDonald A, Hebling A, Wei XP, Yackle K. The breath shape controls intonation of mouse vocalizations. Elife. 2024;13:RP93079. https://doi.org/10.7554/eLife.93079.
43. Maclarnon A, Hewitt G. Increased breathing control: Another factor in the evolution of human language. Evol Anthropol Issues News Rev. 2004;13(5):181–97. https://doi.org/10.1002/evan.20032.
44. Massenet M, Mathevon N, Anikin A, Briefer EF, Fitch WT, Reby D. Nonlinear phenomena in vertebrate vocalizations: mechanisms and communicative functions. Philos Trans R Soc B. 2025;380(1923):20240002. https://doi.org/10.1098/rstb.2024.0002.
45. McKnight JC, Ruesch A, Bennett K, Bronkhorst M, Balfour S, Moss SE, Milne R, Tyack PL, Kainerstorfer JM, Hastie GD. Shining new light on sensory brain activation and physiological measurement in seals using wearable optical technology. Philos Trans R Soc B. 2021;376(1830):20200224. https://doi.org/10.1098/rstb.2020.0224.
46. Molliver ME. Operant Control of Vocal Behavior in the Cat. J Exp Anal Behav. 1963;6(2):197–202. https://doi.org/10.1901/jeab.1963.6-197.
47. Park J, Choi S, Takatoh J, Zhao S, Harrahill A, Han BX, Wang F. Brainstem control of vocalization and its coordination with respiration. Science. 2024;383(6687):eadi8081. https://doi.org/10.1126/science.adi8081
48. Pepperberg IM. Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. Brain Lang. 2010;115(1):81–91. https://doi.org/10.1016/j.bandl.2009.11.002.
49. Perlman M, Clark N. Learned vocal and breathing behavior in an enculturated gorilla. Anim Cogn. 2015;18(5):1165–79. https://doi.org/10.1007/s10071-015-0889-6.
50. Pouw W, Harrison SJ, Esteve-Gibert N, Dixon JA. Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. J Acoust Soc Am. 2020;148(3):1231–47. https://doi.org/10.1121/10.0001730.
51. Ralls K, Fiorelli P, Gish S. Vocalizations and vocal mimicry in captive harbor seals*Phoca vitulina*. Can J Zool. 1985;63(5):1050–6. https://doi.org/10.1139/z85-157.
52. Ravignani A, Fitch WT, Hanke FD, Heinrich T, Hurgitsch B, Kotz SA, Scharff C, Stoeger AS, de Boer B. What Pinnipeds Have to Say about Human Speech, Music, and the Evolution of Rhythm. Front Neurosci. 2016;10:274. https://doi.org/10.3389/fnins.2016.00274.
53. Ravignani A, Kotz SA. Breathing, voice, and synchronized movement. Proc Natl Acad Sci. 2020;117(38):23223–4. https://doi.org/10.1073/pnas.2011402117.
54. Scholander PF. Experimental investigations on the respiratory function in diving mammals and birds. In *Hvalradets skrifter;; Nr. 22*. I kommisjon hos Jacob Dybwad. 1940.

55. Schusterman RJ. Vocal learning in mammals with special emphasis on pinnipeds. In DK. Oller, U Griebel (Eds.), Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication (pp. 41–70). MIT Press: Cambridge, MA. 2008. https://doi.org/10.7551/mitpress/9780262151214.003.0003.
56. Simonyan K, Horwitz B. Laryngeal motor cortex and control of speech in humans. Neuroscientist. 2011;17(2):197–208. https://doi.org/10.1177/1073858410386727.
57. Solomon NP, Garlitz SJ, Milbrath RL. Respiratory and laryngeal contributions to maximum phonation duration. J Voice. 2000;14(3):331–40. https://doi.org/10.1016/s0892-1997(00)80079-x.
58. Stansbury AL, Janik VM. Formant Modification through Vocal Production Learning in Gray Seals. Curr Biol. 2019;29(13):2244–2249 e2244. https://doi.org/10.1016/j.cub.2019.05.071.
59. Stansbury AL, Janik VM. The role of vocal learning in call acquisition of wild grey seal pups. Philos Trans R Soc B. 2021;376(1836):20200251. https://doi.org/10.1098/rstb.2020.0251.
60. Stoeger AS, Mietchen D, Oh S, de Silva S, Herbst CT, Kwon S, Fitch WT. An Asian elephant imitates human speech. Curr Biol. 2012;22(22):2144–8. https://doi.org/10.1016/j.cub.2012.09.022.
61. Sutton D, Larson C, Taylor EM, Lindeman R. Vocalization in rhesus monkeys: Conditionability. Brain Res. 1973;52:225–31. https://doi.org/10.1016/0006-8993(73)90660-4.
62. Titze IR, Martin DW. Principles of voice production. J Acoust Soc Am. 1998;10(1121/1): 424266.
63. Torres Borda L, Jadoul Y, Rasilo H, Salazar Casals A, Ravignani A. Vocal plasticity in harbour seal pups. Philos Trans R Soc B. 2021;376(1840):20200456. https://doi.org/10.1098/rstb.2020.0456.
64. Wiethan F, Ceron MI, Marchetti P, Giacchini V, Mota HB. The use of electroglottography, electromyography, spectrography and ultrasound in speech research-theoretical review. Revista CEFAC. 2015;17:115–25. https://doi.org/10.1590/s1516-18462013005000049.

Publisher's Note