Training Harbor Seals to Participate in Vocal Learning Experiments in a Zoo

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

[1] Comparative Bioacoustics Research Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[2] *Zoo Cleves ("Tiergarten Kleve"), 47533 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

Correspondence: diandra.duengen@mpi.nl

Abstract

Harbor seals (*Phoca vitulina*) are a common zoo species that show a scientifically valuable propensity for vocal learning. However, only a handful of reports exist on their vocal learning abilities and even less document how these were trained or tested. Here, we describe the training and testing of two harbor seals in preparation for vocal learning experiments. We trained the seals to vocalize on visual cues through operant conditioning methods and tested their abilities to i) vocalize and refrain from vocalizing on two distinct cues, and ii) produce two different vocalizations upon the presentation of two individual visual cues. Both seals learned the tasks: the first task was achieved within 118 trials (22 errors to criterion) and 220 trials (40 errors to criterion), the second task within 480 trials (158 errors to criterion) and 380 trials (94 errors to criterion), respectively. We describe a method to train experimentally naive seals for vocal learning experiments. Our results suggest that harbor seals are capable of vocal usage learning and that associating individually distinct vocalizations with different cues may be more cognitively demanding than vocalizing and being silent on cue.

Keywords: Animal training, vocal learning, experiment, pinniped, *Phoca vitulina*, operant conditioning, associative learning

Introduction

Harbor seals are popular in zoos, amongst others for their comparatively easy husbandry, trainability, and willingness to cooperate in shows. Research has shown that (e.g., acoustic) experiments with seals under human care can enrich individuals, thereby improving animal welfare [1]. Harbor seals can be trained to vocalize on a discriminative stimulus (S_D), i.e., a cue [2]. A remarkable case of mammalian plasticity in sound production was a captive male harbor seal, who in the 1980s showed untrained imitation of human speech [3, 4]. A later attempt to replicate vocal imitation in this species failed [Moore [5]]. In his review of vocal learning in pinnipeds, Schusterman [2] for the first time described the fruitful training of vocal variability in a harbor seal. He brought one vocalization under stimulus control and selectively shaped it towards a novel 'wa-wa' sound [2, 6]. Thereby, his case study demonstrates vocal usage learning: the ability to use a vocalization, be it learned or innate, in a novel context as a result of experience [7, 8]. The report provides a valuable starting point to document vocal learning and shaping, one that could spur additional work.

Under human care, vocal usage learning can be tested via operant conditioning experiments [8-13], in which animals are taught to vocalize upon the presentation of an arbitrary cue (e.g., a hand gesture, see Table 1). Each correct response is immediately followed by a reward (e.g., food), whereas incorrect behavior is ignored. Vocal learning via operant conditioning was found in a closely related seal species, the grey seal (*Halichoerus grypus*) [14, 15]. Dissecting vocal learning into its complex features - from 'simple' vocalizing upon the presentation of an S_D to vocalizing and withholding vocalizations upon different S_D's, to producing different vocalizations upon different S_D's - requires a chain of trained behaviors [8, 10, 14].

Building on published work in other species [14, 16], we describe the training and testing process to probe vocal learning in two harbor seals. Our approach combines classical training methods (e.g., operant conditioning, associative learning) and randomized trial orders [17]. We prepared two experimentally naive harbor seals for a series of vocal learning experiments by training them to vocalize, withhold from vocalizing, and emit two different vocalizations upon presentation of respective S_D's [18].

Material & Methods

Setting

The seals were housed in a 230,000-liter freshwater tank and 300 m^2 enclosure at Zoo Cleves, Germany. Two of the six resident harbor seals were selected to participate in this study as they were vocally-active individuals. Seal E was a 6 years old female born at Zoo Osnabrück, Germany. Seal J was a 16 years old male born at Zoo Duisburg, Germany. A third, vocally-active seal dropped out of the experiment due to a severe eye condition. All seals were previously trained to participate in shows, which did not take place during the period of the study. Training was conducted during visitor hours and consisted of general medical training, conducted by zoo staff, and research training, conducted by the experimenter. All *training* took place in the main enclosure, where all seals resided, up to three times/week. Seals were *tested* separately, up to five times per week, in a partitioned area attached to the main enclosure.

Procedure and Training

The seals' behavior was reinforced with pre-cut pieces of mackerel (*Scomber scrombus)* and herring (*Clupea harengus*). In case of a correct trial, the experimenter blew a whistle to indicate the correctness of the behavior, and then reinforced the seal with a randomly chosen piece of fish. In case of an incorrect trial, the experimenter did not blow the whistle, but waited for about 3 seconds (time out).

We distinguish between *Training* and *Testing. Training* sessions were conducted with all seals in one enclosure and constitute a phase where multiple behaviors were trained. These sessions served to establish novel behaviors and maintain known behaviors. Establishing novel behaviors was needed for a series of vocal learning experiments [18] and consisted of e.g., 'stay in a specific spot (station)', 'touch a target with the nose', 'go into the water', and 'follow the trainer'. Much of this training time included habituating the animals to a novel research enclosure, where testing would later take place.

In *Testing*, the seals were tested on dedicated tasks in a separate research enclosure. These tasks consisted of two different experiments, *Experiment 1: Vocalize vs. Silence* and *Experiment 2: Produce Different Vocalizations Upon Different Cues*.

<u>Vocal Training</u>

Seal E and Seal J would occasionally vocalize during *Training* in August and September 2021. We opportunistically reinforced this behavior with fish to increase its occurrence. This training technique is called 'scan and capture' and is common in animal training. Seal E offered one vocalization while Seal J offered two different vocalizations. When the seals produced the vocalizations spontaneously, we reinforced each vocalization and then showed a hand sign to

facilitate association between behavior and cue (see Tab.1, Fig. 1). We assumed the seal would have formed a first association when they responded to approx. 50% of hand signs with the desire behavior (notice here that 50% is not chance level). As a next step, we occasionally reinforced the seals for not vocalizing, i.e., silence. We then trained the seals to associate this silence with a hand sign until they responded correctly to approx. 50% of hand signs. The association of these behaviors and their respective cues were interspersed in the seals' daily training sessions. Only number of training sessions was noted, not the number of each cued behaviors. This opportunistic training approach resulted from the circumstance that at this stage of the study, the seals could not be separated.

Tab. 1: Description of visual cues used to elicit the desired behaviors of the seals

Behavior	**Seal**	**Cue**
Vocalize	E & J	Hand next to the trainer's right ear
Silence	E & J	Stretched arm in front of the trainer's chest with open hand
Vocalization E1	E	Hand next to the right ear (see *Vocalize*)
Vocalization E2	E	Pointed finger rolling in 90-degree angle in front of the trainer
Vocalization J1	J	Right hand and arm above the trainer's head
Vocalization J2	J	Full hand in front of the trainer's mouth with a slight forward movement

Testing

All testing occurred in a separate research enclosure. During *testing*, the loosely formed association between behavior and cue was tested. Trials alternated in a predetermined randomized order, which featured a counterbalanced number of trials and ≤3 trials of the same type in a row [17, 19]. Once the seal was stationed, a trial began and the trainer presented the hand sign for that trial (Fig. 1).

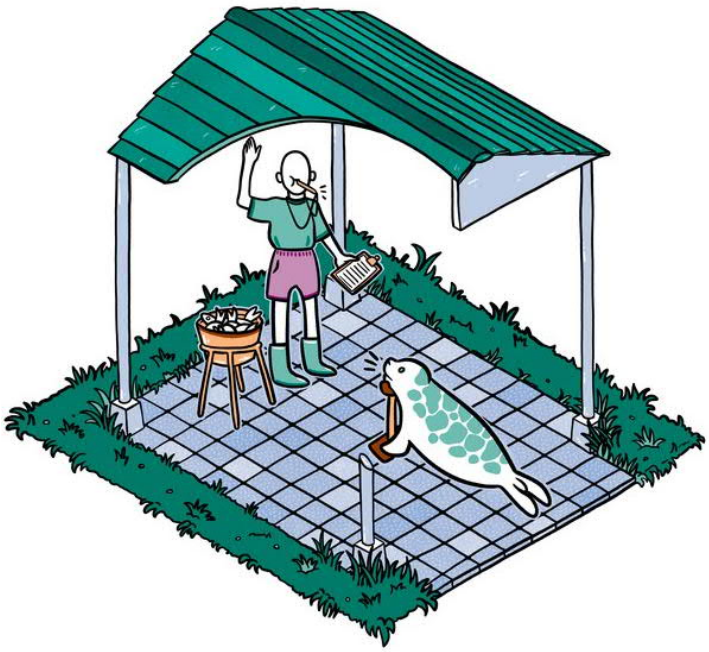

Fig. 1: Illustration of the experimental Setup. The experimenter stands in front of the seal and asks for a pre-determined behavior (here: Vocalization J1, see Tab. 1). Responses were noted on the sheet, which contained the randomized order of behaviors [17, 19].

In rare cases, when an animal was extremely distracted (e.g., due to distraction by zoo visitors) or vocalized incessantly, the trainer repeated the corresponding trial. The decision to repeat a trial was taken immediately by the trainer. Testing was conducted in the months of September and October 2021, and March and April 2022 (Experiment 2, Seal E).

Experiment 1: Vocalize vs. Silence

Sessions consisted of 20-50 trials/session, depending on the animal's motivation; Sessions could be terminated when the animal lacked motivation, e.g. was distracted, chewed on fish rather than swallowing, left the testing area, etc. Trials were counterbalanced between presenting the S_D for vocalizing and withholding from vocalizing. The learning criterion was reached when the animal responded ≥ 80% correctly in four consecutive sessions.

Experiment 2: Produce Different Vocalizations Upon Different Cues

Each session consisted of 20 trials. Trials were counterbalanced and interchanged between asking seal E for vocal type E1/E2 and seal J for J1/J2 (see Fig. 2). The learning criterion was reached when the animal responded ≥ 80% correctly in four consecutive sessions.

Experiment 2 started when seal J successfully completed the *Vocalize and Silence* task on two vocalizations (J1 and J2). Before starting Experiment 2, seal E had to be first trained on a second vocalization. This vocalization, E2, was established by selectively shaping and

reinforcing the initial vocalization (E1) over a period of roughly four months (following the protocol in Schusterman [2]).

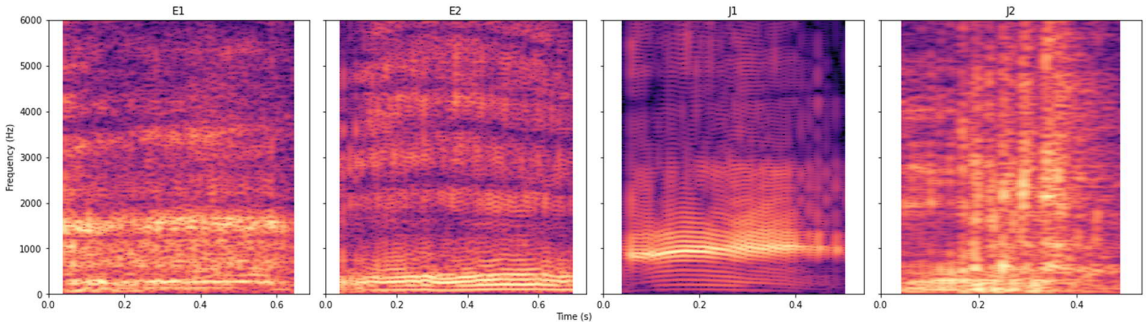

Fig. 2: Spectral representation of the four vocalizations: E1, E2, J1 and J2. Frequency (Hz) on the y-axis, time (s) on the x-axis. Spectrogram was obtained with the Parselmouth package in Python (version 0.4.1, Praat 6.1.38; window length=0.02s, dynamic range=70 dB, max. frequency=7000 Hz [20-22]).

Documentation

During testing, responses were noted on template response sheets (see Fig. 1). A subset of training and testing sessions were also audio and video recorded (Canon Legria HF25 camcorder). Audio recordings were obtained via a Zoom H6 digital recorder connected to a Sennheiser ME-67 unidirectional microphone (frequency response of 40 – 20.000 Hz +- 2.5 dB) covered by a foam windshield.

Data Analysis

We annotated audio recordings in Praat v6.1.38 [20]. Relevant acoustic parameters of vocalizations were then batch-extracted with the Parselmouth package [a Python library for Praat, v0.4.1; 21, 22] in Python. We chose acoustic parameters that were previously analyzed in harbor seal vocalizations, i.e., duration, spectral center of gravity, dominant frequency, percentage voiced, and median harmonicity [23, 24, see Tab. 2].

Tab. 2: Definition of the acoustic parameters and description of the extraction procedure

Name of Acoustic Parameter	**Description of Acoustic Parameter**
Duration (s)	Length of the vocalization in time.
Spectral center of gravity (Hz)	The weighted average frequency of the vocalization's spectrum. The spectrum was calculated with Praat's "Sound: To Spectrum", default parameters. Then, the center of gravity was calculated using "Spectrum: Get center of gravity", default parameters.
Percentage voiced (%)	The pitch was calculated with Praat's "Sound: To Pitch" with default parameters, except pitch floor 50 and pitch ceiling

	400). The number of voiced frames was divided by the number of all frames.
Median harmonicity (dB)	Harmonicity was calculated with Praat's "Sound: To Harmonicity (cc)", default parameters except minimum pitch 50). After this the median of the harmonicity values were calculated.
Dominant frequency (Hz)	The spectrum was calculated with Praat's "Sound: To Spectrum" (default parameters). Then, we selected the frequency bin with the highest power.

Results

Experiment 1: Vocalize vs. Silence

Seal E completed Experiment 1 within 5 sessions i.e., 118 trials (23.6 trials/session ± 4.5) with 22 errors to criterion, the amount of false trials until reaching the learning criterion. Seal J completed Experiment 1 within 6 sessions i.e. 220 trials (36.6 trials/session ± 9.4) with 40 errors to criterion (see Fig. 3,4 and 5). Before testing, Seal E had 9 preceding training sessions, and seal J had 15 preceding training sessions.

Experiment 2: Produce Different Vocalizations Upon Different Cues

Seal E completed Experiment 2 within 480 trials (24 sessions with 20 trials/session) with 158 errors to criterion. Seal J completed Experiment 2 within 380 trials (19 sessions with 20 trials/session) with 94 errors to criterion (see Fig. 3,4 and 5). Four sessions were omitted from analysis as these were terminated early by seal J after a few trials. Before testing, Seal E had 47 preceding training sessions; this high number is due to the extra training she required to establish a novel vocalization type. Seal J had 23 preceding training sessions.

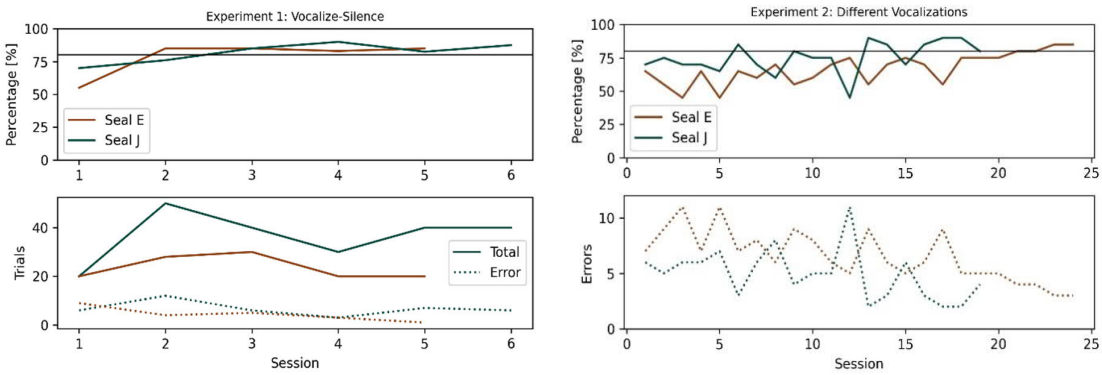

Fig. 3: The upper graphs depict the learning curves of seals E (brown) and J (dark green). The solid line at 80 % on the y-axis marks the learning criterion's threshold. The lower graph shows the seals' errors to criterion (dotted line). Note that the overall trial number per session (solid line) varied for Experiment 1, but was consistent with 20 trials/session in Experiment 2.

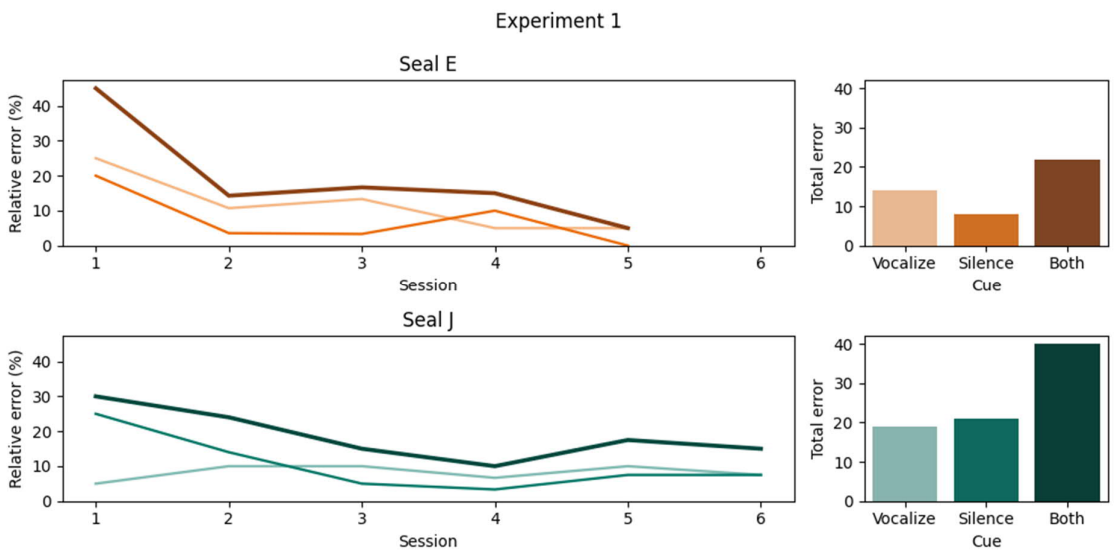

Fig. 4: The lineplot graphs (left, one per seal, seal E: shades of brown, seal J: shades of blue) depict the error to criterion over each session for each cue (Vocalize vs. Silence) separately. The barplots (right, seal E: top, seal J: bottom) show the overall scores for each cue over all sessions.

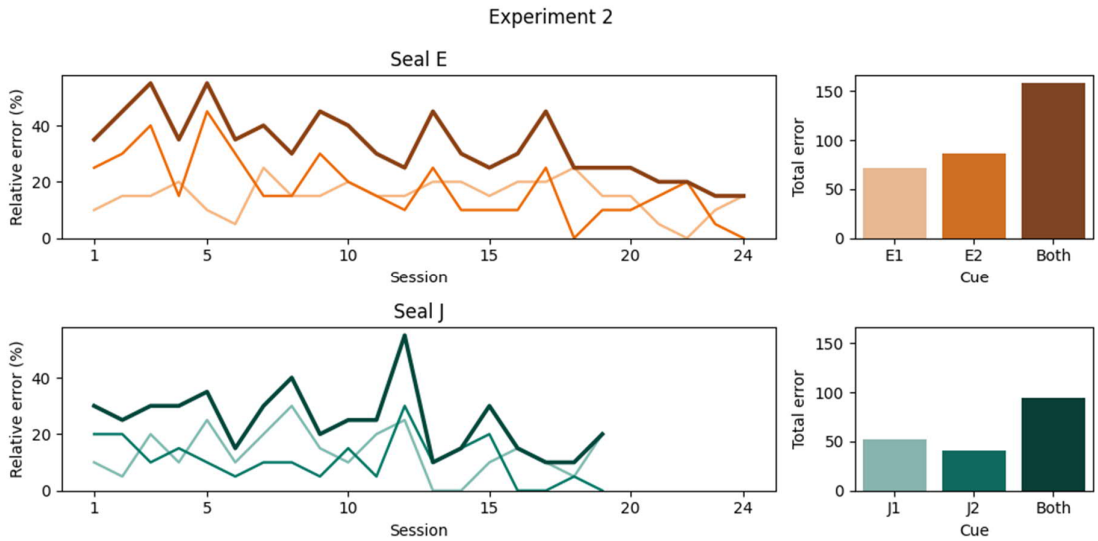

Fig. 5: The lineplot graphs (left, one per seal, seal E: shades of brown, seal J: shades of blue) depict the error to criterion over each session for each cue (Vocalize vs. Silence) separately. The barplots (right, seal E: top, seal J: bottom) show the overall scores for each cue over all sessions.

Discernibility of Vocalization Types

To confirm that the seals produced different vocal types (E1 vs. E2 and J1 vs. J2) and that types were not human subjective categories, we inspected the acoustic parameter composition between call types and within seals (Fig. 6).

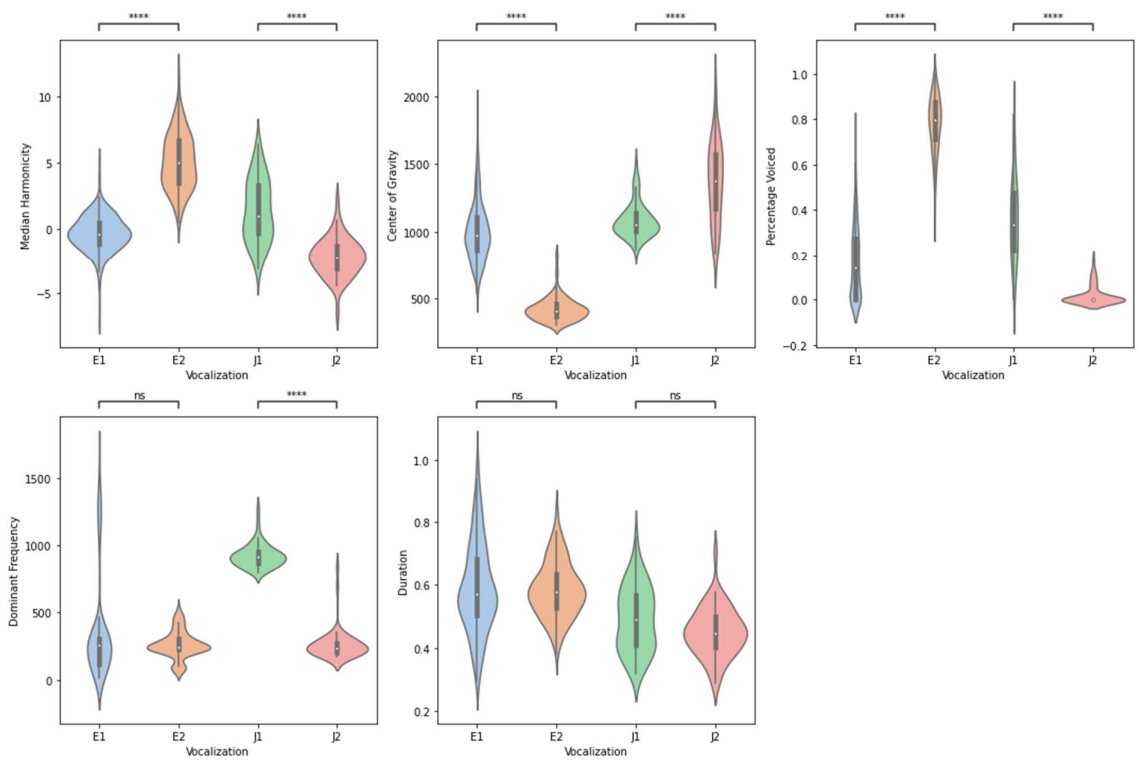

Fig. 6: Visualized spectral parameter distribution of each seal's vocalization types. Significance levels are indicated by asterisks, where ****: $p <= 1.00^{e-04}$, and ns: $5.00e-02 < p <= 1.00^{e+00}$.

A Mann-Whitney-U test revealed that seal E's two vocalization types differed significantly in three of the five extracted spectral parameters, namely center of gravity (statistic=52608.0, p<0.0001), percentage voiced (statistic=171.5, p<0.0001), and median harmonicity (statistic=457.0, p<0.0001). Seal J's two vocalization types differed significantly in four of the five extracted spectral parameters, namely center of gravity (statistic=939.0, p<0.0001), percentage voiced (statistic=4349.5, p<0.0001), median harmonicity (statistic=3825.0, p<0.0001), and dominant frequency (statistic=4470.0, p<0.0001). These significant differences were not affected when correcting for multiple comparisons (five per seal, corresponding to the number of parameters tested).

Discussion

This study provides an operant conditioning protocol to train experimentally naive seals to a vocal learning experiment. In particular, we show how two harbor seals can be trained to associate visual cues with different behaviors, which shows their ability of vocal usage learning.

The task in experiment 1, *Vocalize vs. Silence*, was learned fairly fast, i.e., within 118 and 220 trials (seal E and J respectively). The task in experiment 2, *Produce Different Vocalizations Upon Different Cues*, was learned within 480 trials and 380 trials (seal E and J, respectively). A qualitative comparison between the average trial number of each experiment (*Vocalize and Silence*, $\bar{x}$ = 169 trials; *Produce Different Vocalizations Upon Different cues*, $\bar{x}$ = 430 trials) reveals that the seals needed >2.5 times more trials to learn to *Produce Different Vocalizations Upon Different cues*. Perhaps, experiment 2 was more cognitively demanding than experiment 1; this difference would support the hypothesis that vocal usage learning features components of varying complexity [8, 14, 16], which reflects in the time required to learn tasks.

During experiment 1, the seals were asked to vocalize or remain silent on cue. Both seals succeeded at this task (Fig. 3). Investigating the error to criterion (ETC) value separated per cue revealed that seal J responded almost equally correct (or incorrect) to either of the cues. Seal E, however, showed a preference for vocalizing over silence (Fig. 2). In experiment 2 the seals had to produce different vocalizations upon presentation of different cues. Here, especially seal E showed a much higher ETC value for the E2 vocalization, than for the E1 vocalization (Fig. 4). A reason for this may be the active shaping of the E1 vocalization. This shaping directly preceded experiment 2, causing a temporarily higher exposition of vocalization E1 over E2, which could have led to a type preference. However, seal J showed similarly higher

ETC values for one vocal type (J1) over the other (J2). It is not rare that animals show side or cue preferences [25-27].

In classical cognition tasks, considering the Clever Hans effect is important to avoid falsified results. The Clever Hans effect occurs, when an experimenter involuntary cues an animal. It leads to the animal learning a task based on other cues than intended [28]. Vocal usage learning, however, is independent of this phenomenon, as it is per definition learning to vocalize in association with a certain context [8, 29]. Whether or not the tested animals picked up on intended or unintended cues is irrelevant. Vocal usage learning would be displayed in any case, solely on other cues than anticipated [14].

To study the learning process will require more fine-grained documentation of training sessions from the moment of scanning/capturing the novel behavior. This was unfortunately not possible in this study. Further, a consistent number of trials per session can facilitate optimal comparing learning processes. Following experimental guidelines will further facilitate the direct comparison of vocal learning abilities in different individuals or species [30]. Future studies should address this and track stimulus exposition to better understand underlying (vocal) learning mechanisms.

Overall, our study shows that harbor seals, just like the closely related grey seal [14, 15], can be trained to participate in vocal learning experiments. Experiment 1 shows fast learning of differentiating vocalizing and remaining silent upon demonstration of different visual cues. Experiment 2 shows that the seals are further capable of distinguishing two different vocalization types in response to different S_D's. Future studies should investigate whether harbor seals are also capable of vocal comprehension learning, i.e., understanding vocalizations in specific contexts.

Conclusion

This study aimed at preparing two experimentally naive harbor seals for vocal learning experiments. Our report complements a previous one on vocal shaping and usage learning in a male harbor seal [2]. Here, we show that both female and male harbor seals are capable of vocal usage learning. Our results suggest that this learning implies different levels of complexity, which to the best of our knowledge, has not been shown before in this species. Future studies should incorporate a tracking of the scan/capture progress for a quantitative documentation of the vocal usage learning process. It would be further interesting to build on Shapiro, Slater [14] and Stansbury, de Freitas [15] experimental design on higher levels of usage learning, to see whether harbor seals are capable of such 'more complex' vocal usage learning.

Author Contributions

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

Funding

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.

Institutional Review Board Statement

Ethical review and approval were waived for this study, due to no permit requirement of the general training procedures that were conducted (Az. 81-04.78, Landesamt für Natur, Umwelt und Verbraucherschutz NRW).

Informed Consent Statement

Not applicable

Data Availability Statement

The data can be obtained from the corresponding author upon request.

Acknowledgments

We thank Prof. Dr. Vera Schluessel and Dr. Sara Torres Ortiz for providing helpful comments on the manuscript. We thank Yannick Jadoul for helpful thoughts along the way, the discussions on stimulus randomization, and the brainstorming on plotting. 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 A.R.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Ortiz, S.T., et al., *Does participation in acoustic experiments improve welfare in captive animals? A case study of three grey seals (Halichoerus grypus)*. 2020, Cold Spring Harbor Laboratory.
2. Schusterman, R.J., *Vocal learning in mammals with special emphasis on pinnipeds.* The evolution of communicative flexibility: Complexity, creativity, and adaptability in human and animal communication, 2008: p. 41-70.
3. Ralls, K., P. Fiorelli, and S. Gish, *Vocalizations and vocal mimicry in captive harbor seals, Phoca vitulina.* Canadian Journal of Zoology, 1985. **63**(5): p. 1050-1056.
4. Duengen, D., W.T. Fitch, and A. Ravignani, *Hoover the talking seal.* Current Biology, 2023. **33**(2): p. R50-R52.
5. Moore, B.R., *The evolution of imitative learning*, in *Social Learning in Animals: The Roots of Culture*, C. Heyes and B. Galef Jr, Editors. 1996, Academic Press: 1996:245-265.
6. Goncharova, M., et al., *Vocal tract dynamics shape the formant structure of conditioned vocalizations in a harbor seal.* Annals of the New York Academy of Sciences, 2024.
7. Jarvis, E.D., *Evolution of vocal learning and spoken language.* Science, 2019. **366**(6461): p. 50-54.
8. Janik, V.M. and P.J. Slater, *The different roles of social learning in vocal communication.* Animal Behaviour, 2000. **60**(1): p. 1-11.
9. Adret, P., *Vocal learning induced with operant techniques: an overview.* Netherlands Journal of Zoology, 1993. **43**: p. 125-125.
10. Vernes, S.C., et al., *The multi-dimensional nature of vocal learning.* Philosophical Transactions of the Royal Society B, 2021. **376**(1836): p. 20200236.
11. Stoeger, A.S. and A. Baotic, *Operant control and call usage learning in African elephants.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2021. **376**(1836): p. 20200254.
12. Seyfarth, R.M. and D.L. Cheney, *Production, usage, and comprehension in animal vocalizations.* Brain and language, 2010. **115**(1): p. 92-100.
13. Lattenkamp, E.Z., S.C. Vernes, and L. Wiegrebe, *Volitional control of social vocalisations and vocal usage learning in bats.* J Exp Biol, 2018. **221**(Pt 14).
14. Shapiro, A.D., P.J. Slater, and V.M. Janik, *Call usage learning in gray seals (Halichoerus grypus).* Journal of Comparative Psychology, 2004. **118**(4): p. 447-54.
15. Stansbury, A.L., et al., *Can a gray seal (Halichoerus grypus) generalize call classes?* Journal of Comparative Psychology, 2015. **129**(4): p. 412.

16. Vergara, V. and L. Barrett-Lennard, *Call Usage Learning by a Beluga (Delphinapterus leucas) in a Categorical Matching Task.* International Journal of Comparative Psychology, 2017. **30**.
17. Gellermann, L.W., *Chance orders of alternating stimuli in visual discrimination experiments.* The journal of genetic psychology, 1933. **42**: p. 206-208.
18. Duengen, D., Jadoul, Y., Ravignani, A., *Vocal usage learning and vocal comprehension learning in harbor seals.* under review.
19. Jadoul, Y., D. Duengen, and A. Ravignani, *PyGellermann: a Python tool to generate pseudorandom series for human and non-human animal behavioural experiments.* BMC Research Notes, 2023. **16**(1): p. 135.
20. Boersma, P. and D. Weenink, *Praat: doing phonetics by computer* in *Glot International*. 2022.
21. Jadoul, Y., B. Thompson, and B. De Boer, *Introducing parselmouth: A python interface to praat.* Journal of Phonetics, 2018. **71**: p. 1-15.
22. Jadoul, Y., B. De Boer, and A. Ravignani, *Parselmouth for bioacoustics: automated acoustic analysis in Python.* Bioacoustics, 2023: p. 1-17.
23. Van Parijs, S.M., G.D. Hastie, and P.M. Thompson, *Individual and geographical variation in display behaviour of male harbour seals in Scotland.* Animal Behaviour, 2000. **59**: p. 559-568.
24. Raimondi, T., Haas, C.E., et al., *Less variable auditory input increases vocal individuality in harbour seal pups (Phoca vitulina).* under review.
25. Ortiz, S.T., et al., *Problem solving capabilities of peach-fronted conures (Eupsittula aurea) studied with the string-pulling test.* Behaviour, 2019. **156**(5-8): p. 815-846.
26. Erdsack, N., G. Dehnhardt, and F.D. Hanke, *Serial visual reversal learning in harbor seals (Phoca vitulina).* Animal Cognition, 2022: p. 1-11.
27. Schluessel, V., H. Kraniotakes, and H. Bleckmann, *Visual discrimination of rotated 3D objects in Malawi cichlids (Pseudotropheus sp.): a first indication for form constancy in fishes.* Animal cognition, 2014. **17**: p. 359-371.
28. Sebeok, T.A. and R.E. Rosenthal, *The Clever Hans phenomenon: Communication with horses, whales, apes, and people.* Annals of the New York Academy of Sciences, 1981.
29. Janik, V.M. and P.J.B. Slater, *Vocal learning in mammals*, in *Advances in the Study of Behavior, Vol 26*, P.J.B. Slater, et al., Editors. 1997, Elsevier Academic Press Inc: San Diego. p. 59-99.
30. Lattenkamp, E.Z., et al., *A researcher's guide to the comparative assessment of vocal production learning*. Philosophical Transactions of the Royal Society B, 2021. **376**(1836): p. 20200237.