DOI: 10.1111/nyas.15189

ORIGINAL ARTICLE

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Vocal tract dynamics shape the formant structure of conditioned vocalizations in a harbor seal

Maria Goncharova[1] | Yannick Jadoul[1,2,3] | Colleen Reichmuth[4] | W. Tecumseh Fitch[5,#] | Andrea Ravignani[1,3,6,#]

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

[2]Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium

[3]Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

[4]Long Marine Laboratory, Institute of Marine Sciences, University of California Santa Cruz, Santa Cruz, California, USA

[5]Department of Behavioral and Cognitive Biology, Vienna CogSciHub, University of Vienna, Vienna, Austria

[6]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

Correspondence
Maria Goncharova, Comparative Bioacoustics Research Group, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525XD, the Netherlands. Email: maria.goncharova@mpi.nl

Andrea Ravignani, Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy. Email: andrea.ravignani@uniroma1.it

Colleen Reichmuth, Long Marine Laboratory, Institute of Marine Sciences, University of California Santa Cruz, Santa Cruz, California, USA. Email: coll@ucsc.edu

W. Tecumseh Fitch, Department of Behavioral and Cognitive Biology, Vienna CogSciHub, University of Vienna, Vienna, Austria. Email: tecumseh.fitch@univie.ac.at

[#]W. Tecumseh Fitch and Andrea Ravignani are co-senior authors.

Funding information
European Research Council, Grant/Award Number: Advanced Grant SOMACCA; Austrian Science Foundation Grant, Grant/Award Number: (#W1262-B29); Danmarks Grundforskningsfond, Grant/Award Number: DNRF117; Office of Naval Research, Grant/Award Number: N00014-04-1-0284; Max-Planck-Gesellschaft, Grant/Award Number: Independent Max Planck Research Group Leader funding

Abstract

Formants, or resonance frequencies of the upper vocal tract, are an essential part of acoustic communication. Articulatory gestures—such as jaw, tongue, lip, and soft palate movements—shape formant structure in human vocalizations, but little is known about how nonhuman mammals use those gestures to modify formant frequencies. Here, we report a case study with an adult male harbor seal trained to produce an arbitrary vocalization composed of multiple repetitions of the sound *wa*. We analyzed jaw movements frame-by-frame and matched them to the tracked formant modulation in the corresponding vocalizations. We found that the jaw opening angle was strongly correlated with the first (F1) and, to a lesser degree, with the second formant (F2). F2 variation was better explained by the jaw angle opening when the seal was lying on his back rather than on the belly, which might derive from soft tissue displacement due to gravity. These results show that harbor seals share some common articulatory traits with humans, where the F1 depends more on the jaw position than F2. We propose further in vivo investigations of seals to further test the role of the tongue on formant modulation in mammalian sound production.

KEYWORDS
articulation, formants, *Phoca vitulina*, source-filter theory, vocal communication, vocal tract

INTRODUCTION

According to the source-filter theory of sound production,[1,2] animal vocalizations can be described by the specific resonance frequencies of their upper vocal tract cavities, which filter the sound produced by the larynx or syrinx (i.e., source). The resonance frequencies of the supra-laryngeal tract selectively enhance or dampen specific regions of the frequency spectra of the sounds that originate from the larynx; the resulting peaks in this acoustic filter's spectrum affect the spectral envelope of the vocalization. These resonance frequencies are called formants and broadly shape the spectral envelope of the resulting vocalization.

Formants play an essential role in the vocal communication of a broad range of animal taxa, including mammals,[3–5] birds,[6] and reptiles.[7] Formants reliably reflect the conformation of the upper vocal tract, which differs between individuals and, as a rule, correlates strongly with body size.[8] Therefore, formants might encode information about the caller's identity, sex, physical condition, as well as hormonal and emotional state.[4,6,9–11]

Humans are the mammalian species where formants are most studied and best understood. From an acoustic perspective, sounds of human speech are vocal signals with distinctive formant structure; only minor variations in this structure are allowed without disturbance to the intelligibility of a speaker.[12,13] Humans fine-tune the formant structure of speech sounds using articulatory gestures including movements of the jaws, tongue, lips, and soft palate. These movements dynamically change the volume and shape of cavities that form the upper vocal tract, which in turn causes changes in the formant structure of the resulting sound.

Despite the fact that human phonation mechanisms are understood in considerable detail, it is not possible to generalize human-based vocal emission models to all mammals. The mammalian upper vocal tract may have been shaped by multiple, potentially competing, evolutionary pressures,[14] as this structure is not only used to produce sounds but is also involved in feeding, breathing, and thermoregulation. If the upper vocal tract evolves and transforms to support one of its many functions, this may also influence the other functions that use the same anatomical structures. The use of human-inspired principles of dynamic formant control has been shown in a few terrestrial mammals.[15–19] However, some principles of vocal production and emission are drastically different in cetaceans, which are secondarily aquatic mammals.[20,21] Therefore, mapping the evolution of mammalian articulatory abilities requires examining similarities and differences in articulation across mammals to reveal the underlying physical and physiological mechanisms, particularly in other aquatic mammals (e.g., pinnipeds or sirenians).

Many mammalian species are capable of temporary elongation of the upper vocal tract via active larynx lowering.[16,22–27] Vocal tract elongation exaggerates the acoustically apparent size of the caller (i.e., the size perceptually inferred from sound by a receiver is larger than the actual body size of the caller). Other mammals use mandible lowering and lip protrusion for modulation of formant structure in the process of a single sound emission.[15,17–19,28] Little is known about mechanisms underlying mammalian phonation beyond individual sounds; in particular, whether and for which vocalizations rapid upper vocal tract movements may cause the sequential formant variation remains untested across mammals. Although such upper vocal tract movements can easily be observed in many common species (e.g., dogs, cats, cattle, etc.), this hypothesis has only been directly tested in humans[29] and remains unexplored in most other mammals. Understanding such a dynamic mechanism is key to mapping similarities and differences in vocal production across mammals, including humans.

Phocid seals can serve as potential model subjects for investigating the mechanisms of articulation in mammals. As diver–hunters, phocids have outstanding control over their respiratory system,[30] potentially allowing them to emit underwater vocalizations up to 1.5 min long.[31] Some phocids appear to have a high degree of vocal flexibility, and a few may be vocal learners (i.e., possess a rare ability to learn new vocal signals).[32,33] Some seals can learn to produce sounds through training, which allows exploration of the abilities of a vocal apparatus both within and beyond their natural repertoire.[34] Properly trained phocids can both start and stop sound production on a trainer's cue.[35] This ability provides an opportunity to control the duration of artificial vocalizations taught to a participating animal in a broad range, which potentially allows a researcher to observe the phonation mechanisms of sound emission for both short and long durations.

In the present study, we recorded the vocalizations and mandible movements of an adult male harbor seal that was reared in captivity and trained to produce airborne vocalizations on cue. Here, we focus on the learned *wawa* vocalization, a series of repetitions of the sound similar to the English syllable *wa*. This unusual sound has a formant structure that modulates in a periodic manner, with alternations of sounds roughly corresponding to the /*w*/ and /*a*/ sounds in English.[36]

Based on initial auditory and visual impressions, we hypothesized that harbor seals are capable of dynamically modulating the formant structure of their vocalizations. While we did not conduct quantitative modeling of the vocal tract, we predicted that these articulatory abilities would follow general mammalian (including human) biomechanics of the upper vocal tract as previously clarified for humans. More precisely, we predicted that just like in humans and human-based models: (1) mandible lowering and a corresponding increase of jaw gape is associated with F1 and F2 rise, and (2) mandible lowering has a larger effect on F1 than F2.

METHODS

Subject and vocal training

To investigate the link between vocal tract shape and vocal production in phocids, we analyzed audio and video data streams collected at Long Marine Laboratory at the University of California, Santa Cruz in February 2007. Our subject was a mature male harbor seal identified as Sprouts (NOA0001707). This seal was trained using operant conditioning methods and positive food reinforcement to perform various

husbandry and research tasks during his lifetime.[37,38] At the time of this study, Sprouts was 18 years old and weighed 108 kg. He produced few spontaneous sounds in the air, with occasional snores, splutters, and noisy nasal exhalation sounds. However, he had been previously conditioned to emit several distinctive airborne vocalizations on cue.[34]

One of these sounds, which we termed the *wawa* vocalization, was a series of a variable number of concatenated *wa* sounds. Initially, the seal had been rewarded for producing any vocalization, after which he was trained to produce variable specific vocalizations. From these, the syllable-like sound *wa* was selectively reinforced and then shaped to be emitted in long, repetitive sequences.[34] Production of the *wawa* vocalization involved visible jaw oscillations with each *wa* sound (Supporting information S1). Here, we focus on documenting these *wawa* sounds, which allowed us to best observe and quantify the relationship between formant modulations and jaw movements.

Animal research was conducted without harm under the authorization of NOAA/NMFS marine mammal research permit 1072–1771 in accordance with the animal welfare laws of the United States. This study was approved by the Institutional Animal Care and Use Committee at the University of California, Santa Cruz.

Data collection

Acoustic recordings were conducted in a sound-attenuating chamber, which minimized background noise and reverberation.[39] Sprouts voluntarily entered the chamber with his trainer (C.R.) and several assistants. During a single session on the day of the recording, his movements and sound production were documented with audio and video recordings; corresponding ultrasound data (not considered here) were collected from a handheld probe gently placed on his throat. We placed three white zinc-oxide marks on his black fur to highlight standard landmarks for video analysis. The vibrissal point (V) was on the side of the snout, just behind the posterior-most vibrissae; the angular point (A) was at the posterior corner of the mouth, and the jaw point (J) was on the mandible about 3 cm posterior to the chin (Figure 1). The trainer prompted Sprouts to produce the *wawa* vocalization. His vocal response was intermittently reinforced with a conditioned stimulus (whistle) followed by several pieces of fish. Under these conditions, the seal was highly cooperative and compliant.

Audio was captured from <1 m distance with a Neumann 82i condenser shotgun microphone linked to a Marantz PMD-660 solid-state recorder. Audio was recorded with a 48 kHz sampling rate and 16-bit quantization. Corresponding video data were obtained with a Sony DCR-PC9 NTSC camera mounted on a small tripod placed 90 degrees to the seal's midline, about 40 cm away. The video was recorded in SP mode (highest quality; 29.97 FPS) and also captured sound. Audio and video data streams were later aligned based on the maximum cross-correlation of both audio tracks.

One set of audio-video recordings was obtained with the seal in the prone position, resting on his belly, and another set was obtained with him lying on his back in the supine position (Figure 1). The seal vocalized readily in both configurations. Since harbor seals normally vocalize under water with flexible body orientation,[40,41] neither position was essentially more ecologically valid than the other. Thus, we can assume that the animal was not forced to vocalize in a physiologically implausible body orientation during the data collection.

Audio analysis

Audio signals were stored as uncompressed *.wav* files. We examined the spectrograms and annotated every sound on the recordings (e.g., Sprouts' vocalizations, human speech, and other noises) using the software package Praat. We used the annotated *wawa* vocalizations to investigate the variability of their duration and number of *wa* sounds in each vocalization.

After that, a custom Python script (using the *TextGridTools* package[42]) served to find the time coordinates of all the intervals that simultaneously contained *wawa* vocalizations without any overlaps with other sounds. The same script then selected nonoverlapping intervals where Praat's intensity curve (default parameters) of the audio exceeded a threshold of 25 dB below the maximum intensity of that particular audio file.

The frequency values of the first two formants (F1 and F2) were identified within selected intervals using a custom Python script featuring the Parselmouth library (a Python interface to Praat, version 0.4.0, Praat version 6.1.38).[43,44] After filtering out frequencies above 1500 Hz (Praat's *Filter (stop Hann band)*, 50 Hz smoothing), the script extracted formant estimates calculated via Praat's automated *To Formant: Burg* analysis within the time intervals selected during the previous step of the analysis (i.e., a *wawa* vocalization fragment with no silent gaps or overlaps with any other sounds). Praat's analysis estimated three formants below a maximum frequency of 3000 Hz with an analysis window length of 0.05 s; for all other parameters, we kept Praat's default values. We then linearly interpolated the first two formants (F1 and F2) at the time point of the start of each new video frame ensuring the extracted F1 and F2 values directly corresponded to the measurements of the jaw position described below. We did not extract the third formant (F3), as we judged the quality of the automatic tracking to be insufficient. In summary, for each audio interval, the analysis resulted in a multivariate time series of the first two formant estimates sampled to match the video frame rate (i.e., 29.97 FPS, or approximately every 33 ms).

Video analysis

We extracted the video fragment corresponding to each audio interval selected for analysis and tracked the serial kinematics of jaw movement. To do this, we extracted each video frame as a *png* image using a custom Python script (featuring the *imageio*, *PIL*, and *NumPy* packages).[45–47] These still frames were evaluated using ImageJ software[47] with the *Manual Tracking* plugin to trace the coordinates of the vibrissal, angular, and jaw points (V, A, and J) on the seal (see

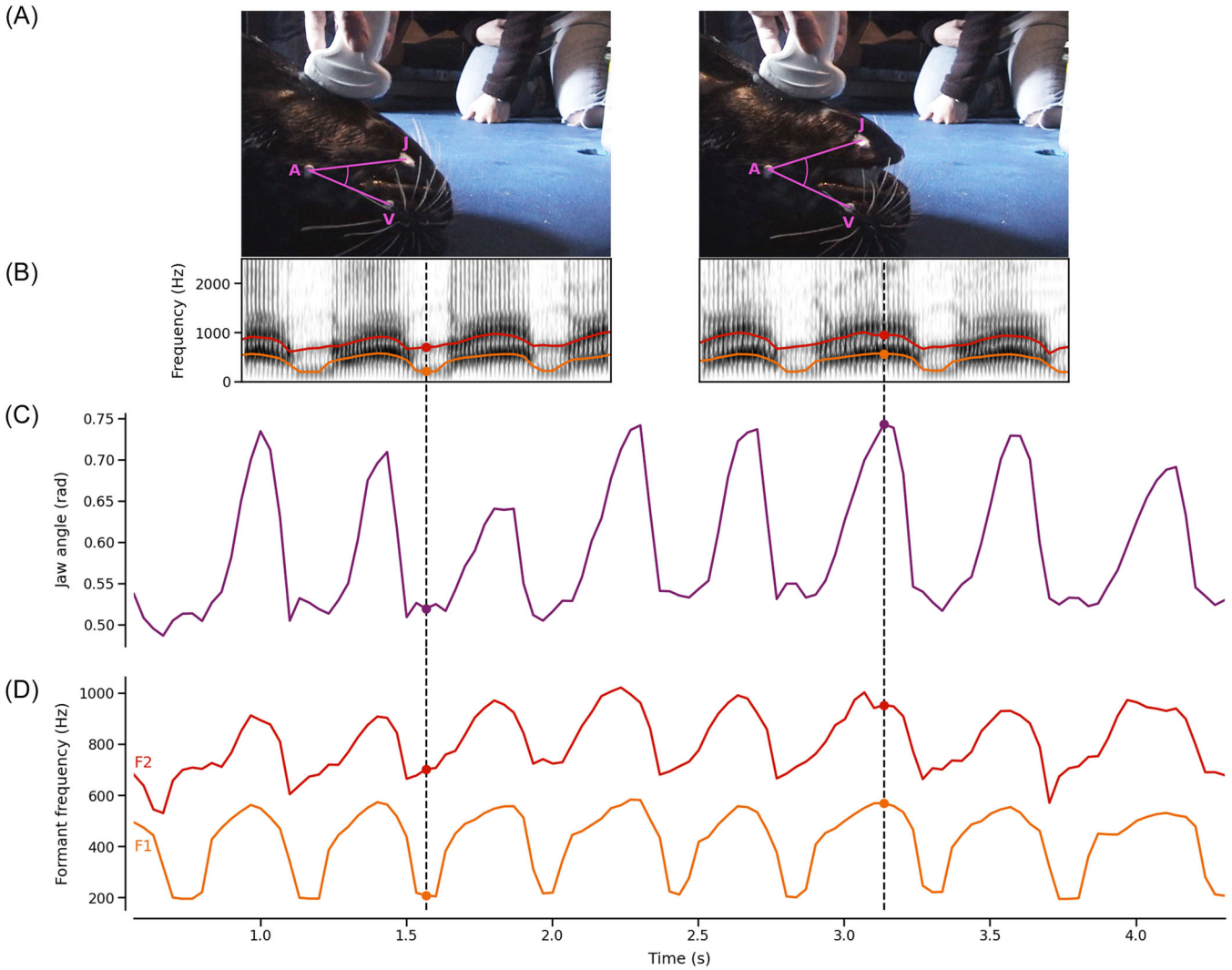

FIGURE 1 (A) Two exemplary contrast-enhanced video frames show the landmarks used for the jaw movement tracking (A, angular point; V, vibrissal point; J, jaw point) and the resulting jaw opening angle at selected points in time (dashed lines). (B) Spectrograms of the recorded audio show the clear formant structure and the automatically extracted formant tracks. (C) The change in jaw angle over time closely matches the F1 and F2 modulations (D), as illustrated here by an example of a *wawa* vocalization that the harbor seal Sprouts produced in the supine position.

Figure 1). We calculated two variables from these coordinates: jaw gape and jaw angle. Jaw gape was measured as the distance in pixels between the V and J points. Jaw angle was computed as the angle between the $\overrightarrow{AV}$ and $\overrightarrow{AJ}$ vectors, measured in radians. We also determined the jaw oscillation rate per second for each separate *wawa* vocalization by dividing the total duration of the vocalization by the number of *wa* sounds in the vocalization.

Statistical analysis

The previous steps resulted in a dataset containing the combined numeric data from the parallel, time-synchronized audio and video streams. This dataset contained, for each selected and analyzed interval, the values of F1 and F2 for each time point, the geometric descriptors of lower jaw movements (jaw gape and jaw angle), and the position of the seal (prone versus supine). We provide this dataset as Supporting Information S2. Since jaw gape and jaw angle are geometrically related measurements, these two variables were strongly correlated in both positions (Pearson correlation: $r = 0.78$, $p < 0.001$ in the prone position, $r = 0.91$, $p < 0.001$ in the supine position). Hence, to avoid introducing both correlated independent variables into the same statistical model, we only used the jaw angle in further analyses. We picked the jaw angle over the jaw gape for two reasons: The jaw angle's measurement unit (radians) is easier to interpret than that of the jaw gape (i.e., pixels, lacking an absolute external reference in the videos), and the jaw is easier to compare between studies as it is not tied to the overall linear size of the animal.

We performed a series of Spearman correlations to investigate the correlation between the jaw angle and F1/F2. Nonparametric methods were used as most relevant variables were not normally distributed (Kolmogorov–Smirnov test, $p < 0.05$).

We could not be sure that Sprouts produced the *wawa* vocalizations in the exact same manner each time (i.e., minor differences in articulation could be caused by differences in arousal level, exact positioning, or other confounding behavioral factors). Moreover, we had to consider the artificial discontinuities that we introduced by our method of interval segmentation and selection. As pooling all data points in a single nonparametric statistical test would fail to take into account these concerns, we performed separate statistical tests for each interval. A meta-statistic examination of the test results was conducted to detect the general trends underlying the sound production we observed. We calculated the number and percentage of tests that detected a significant correlation between two given variables (jaw angle versus F1 or jaw angle versus F2) for the subset of vocalization intervals for each body position (supine or prone position) and for all intervals pooled. We also calculated the median, interquartile range, and min-max of Spearman rank r for each subset of tests within body position and for all tests.

Finally, we investigated the differences in median jaw angle, F1, F2, duration of *wawa* vocalizations, number of *wa* sounds in a vocalization, and jaw oscillation rate between different body positions with six Kruskal–Wallis tests.

RESULTS

We analyzed 39 *wawa* intervals produced by the seal, 19 for the prone position and 20 for the supine position. In every case, the formant modulation pattern closely matched the jaw oscillation pattern (Figure 1). F1, F2, and jaw angle significantly differed between body positions: Sprouts produced the first two formants with lower frequencies when supine than when prone. In the supine position, the jaws moved within a smaller angular range (Figure 2).

The duration and number of *wa* sounds in *wawa* calls did not vary with position, while the jaw oscillation rate was greater when the seal was prone. We found no significant differences between body positions in any of these variables, except for the jaw oscillation frequency (Kruskal–Wallis test, Figure 2).

Regardless of body position, the first two formant values strongly correlated with the jaw angle (jaw angle versus F1: Spearman rank $r = 0.60$, $p < 0.001$; jaw angle versus F2: $r = 0.31$, $p < 0.001$, Figure 3). The strength of the correlations between jaw angle and F1 and F2 and the proportion of intervals that showed correlations varied between body positions. When the intervals were considered separately, the significant correlation between jaw angle and F1 was evident in 38/39 cases (Spearman rank correlations r, $p < 0.001$; Table 1). F2 was significantly correlated with jaw angle in 27/39 cases. The correlation between F2 and jaw angle was more robust for the supine position (19/20 intervals) than for the prone position (8/19 intervals) (Fisher's exact test $p < 0.001$; Table 1).

For all intervals, jaw angle was positively correlated with F1 (all Spearman rank $r \geq 0.13$; Table 1) (i.e., the wider Sprouts opened his jaws, the higher the F1 frequency of his sounds). We observed the same correlation pattern between jaw angle and F2 in the supine but not in the prone position. This further supports the finding that the correlation between jaw angle and F2 is more robust in the supine position (Table 1).

DISCUSSION

Effect of jaw angle on vocal output

Regardless of body position, the gape of the seal's mouth was related to the frequencies of vocal formants. The jaw opening angle was positively correlated with F1 in nearly all intervals tested and with F2 in two-thirds of the intervals tested. Jaw angle had a bigger effect on F1 than on F2, and this effect was more prominent in the prone position. These results correspond to available human data and the theoretical fundamentals of the physics of speech, thus supporting our initial hypothesis.[1,48]

In the supine position, the correlation results were similar for both F1 and F2. We found significant positive correlations between jaw angle and F1 and F2 in almost all instances. We hypothesize that the increased susceptibility of F2 to jaw angle in the supine position might derive from changes in tongue position and soft tissue displacement due to gravity. For instance, in humans, the pharyngeal cavity tends to be smaller in the supine than in the upright position, which might result from the tongue root shifting posteriorly.[49] However, further investigation with in vivo visualization would be required to test this hypothesis for seals.

We did not formally analyze F3, but visual inspection of spectrograms suggests that F3 is mostly constant and not correlated with jaw movements. This again matches human data and existing models where such a correlation is typically absent.[50]

Effect of body position

The seal's body position had a substantial effect not only on the correlation between jaw angle and F2 but also on the absolute values of the first two formants, the measured angle of jaw opening, and jaw oscillation frequency. Again, we suggest that these differences might be caused by changes in the relative direction of gravity. The above-mentioned changes might directly affect the placement and movement patterns of the mandible and the tongue, along with the deformation of the soft tissues of the upper vocal tract cavities. This, in turn, might have led to the differences in the resonance frequencies of the upper vocal tract between the body positions. The effects of body position on the dimensions of the upper vocal tract, and in particular, on the properties of the tongue and the characteristics of mandible movements are well-studied in humans[51–56] but not in other mammals. It has been shown that positional changes in humans alter the volumes of the upper vocal tract cavities but not the vocal tract length.[49]

When it comes to particular sounds of speech (for instance, vowels in English), the reports are contradictory. With a change of body position from prone to supine, the first two formant values are either reported to be stable[49] or to rise in the case of F1 and decrease in the

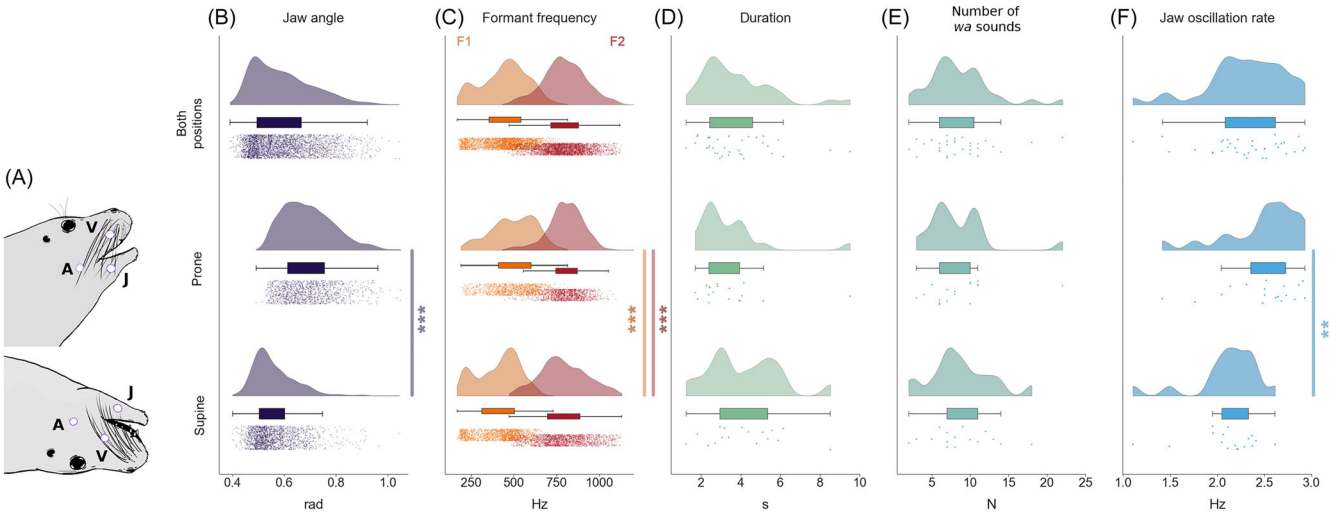

FIGURE 2 Range and distribution of the values of jaw angle, first two formant frequencies, duration, number of syllable-like *wa* sounds, and jaw oscillation rate for the seal's artificial *wawa* vocalizations are shown for prone, supine, or regardless of body position. (A) Location of the reference points used for video analysis of the seal's head in prone and supine positions: V, vibrissal point; A, angular point; J, jaw point. For illustrative and analytic purposes, we provide a schematic outline of Sprouts' head. (B) shows the scatter of raw data points (*rain*), the distribution of raw data points (*clouds*), the median and the quartiles (box), and the min-max range (whiskers) for the jaw angle (measured in radians) for both positions pooled ($n = 39$ intervals) and for the prone ($n = 19$ intervals) and supine positions separately ($n = 20$ intervals). Kruskal–Wallis test shows a difference in jaw angle ($H_{1,3031} = 1657.44$, $p < 0.001$) and the first two formant values (F1: $H_{1,3031} = 294.65$, $p < 0.001$; F2: $H_{1,3031} = 16.59$, $p < 0.001$) between the two body positions. (C) shows the same descriptors of the same data sample as panel B for the first two formant frequencies (F1 and F2). (D) shows the scatter of raw data points (*rain*), the distribution of raw data points (*clouds*), the median and the quartiles (box), and the min-max range (whiskers) for the duration of *wawa* vocalizations for both positions pooled ($n = 17$ complete *wawa* vocalizations), and for the prone ($n = 9$ *wawa* vocalizations) and supine positions separately ($n = 8$ *wawa* vocalizations). (E) shows the same descriptors of the same data sample as panel D for the number of syllable-like *wa* sounds in a *wawa* vocalization. (F) shows the same data as the panel D for the frequency of jaw oscillations in a *wawa* vocalization. The Kruskal–Wallis test shows a difference in jaw oscillation rate between two body positions ($H_{1,35} = 9.02$, $p < 0.01$). Significant differences between body positions are indicated (Kruskal–Wallis test, $^{**}p < 0.01$, $^{***}p < 0.001$).

TABLE 1 Summary of Spearman rank correlations between jaw angle and the first two formant variables performed separately for each *wawa* vocalization.

	Jaw angle					
	F1			F2		
	Both positions	Prone	Supine	Both positions	Prone	Supine
N_{sig} / N_{total}	38 / 39	18 / 19	20 / 20	27 / 39	8 / 19	19 / 20
Percentage of N_{sig}	97.44%	94.74%	100%	69.23%	42.11%	95%
Spearman rank *r*	0.70	0.75	0.66	0.39	0.13	0.65
[Q25; Q75]	[0.61; 0.80]	[0.67; 0.84]	[0.58; 0.75]	[0.06; 0.73]	[−0.11; 0.42]	[0.54; 0.81]
(min; max)	(0.33; 0.92)	(0.48; 0.92)	(0.33; 0.92)	(−0.54; 0.94)	(−0.54; 0.85)	(0.17; 0.94)

Note: The proportion and percentage of tests that detected a significant correlation to the total sample of intervals analyzed in a given positional category. Spearman correlation *r*, median, [Q25; Q75], and (min; max) summary values are provided. N_{sig} stands for the number of Spearman correlation tests that did reveal a significant correlation ($p < 0.05$), whereas N_{total} stands for the total number of tests made for the given set of conditions (Prone, Supine, or both positions combined).

case of F2.[55] Therefore, we suggest that the momentary articulatory state of the vocal tract in harbor seals might interact with the effect of the body position and thus contribute to vocal output in various ways.

In the present study, both F1 and F2 were consistently lower for the supine than for the prone body position. The discrepancy between our results and those found in human studies might be a result of at least two factors: (1) Despite the overall similarity between the seal and human vocal tracts,[57] humans and pinnipeds differ in skull and neck anatomy so the effect of body position on the volume of upper vocal tract chambers and the length of the vocal tract in seals might vary from the effects demonstrated in humans; (2) humans might fine-tune

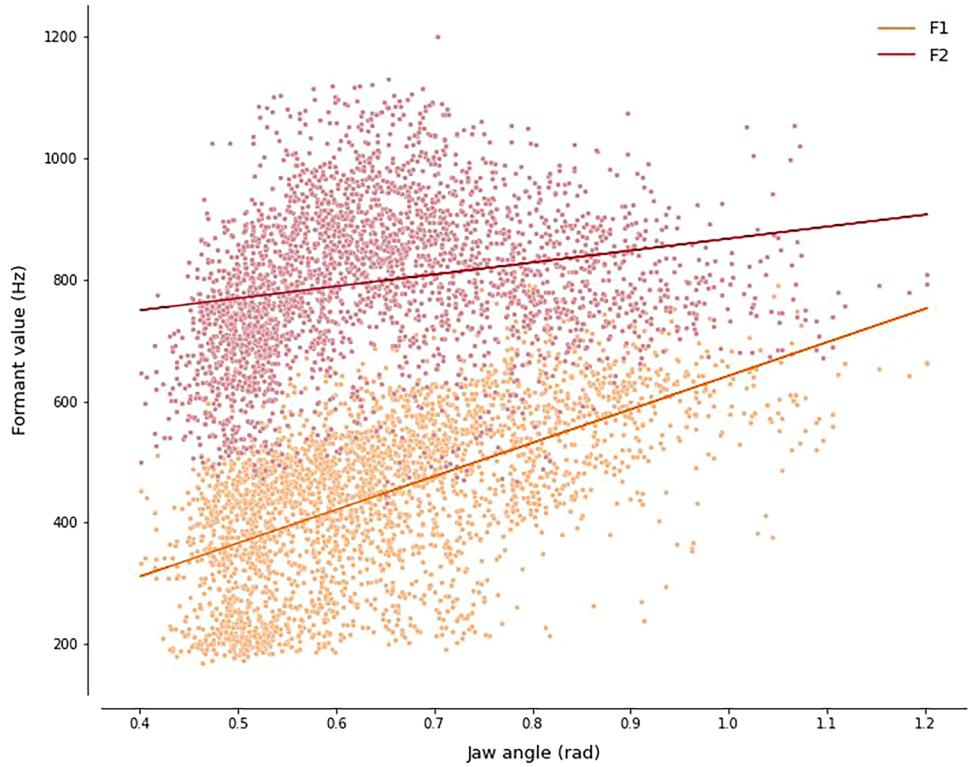

FIGURE 3 Values of the first two formants of Sprouts' *wawa* vocalization strongly correlated with jaw angle (jaw angle versus F1: Spearman rank $r = 0.60$, $p < 0.001$; jaw angle versus F2: $r = 0.31$, $p < 0.001$). For illustrative purposes, linear least-squares regression lines are shown for F1 and F2.

their articulatory gestures in order to actively compensate for the arbitrary disturbances caused by body position and preserve the formant structure of a given vocalization. Unfortunately, the relationships between body position, articulation, and formant modulation in other mammal species are poorly understood because of the few studies covering this topic. Thus, we believe that our findings make a valuable initial contribution to the investigation of potentially universal mechanisms of mammalian vocal production.

CONCLUSIONS AND FUTURE PROSPECTS

In this study, we demonstrate that harbor seals are capable of performing some of the patterns of articulatory-dependent formant modulation that have been described for humans. The jaw opening angle in the seal caused a rise in F1 regardless of body position. The absolute values of both F1 and F2 and the jaw position and oscillation frequency values varied between prone and supine body positions. At the same time, the influence of jaw opening angle on F2 was, in general, weaker than on F1; this makes sense because F2, more than F1, is more related to other articulators, such as lip rounding and tongue position in humans.[1,48,58] One fascinating aspect of our results is the highly dynamic manner in which formants are modified. We know of no previous evidence for such a highly reliable and repeatable variation of formants during a single continuous vocal exhalation in a nonhuman animal. This pattern is temporally consistent and reminiscent of the syllable repetition seen in infants' and bat pups' reduplicated babbling.[59] Although we do not claim that the sound patterns seen in our seal are directly comparable to human syllables, future work might explore testing other articulatory mechanisms that underlie sound production in both species.

Future in vivo imaging of seal articulatory abilities should consider dynamic alterations in tongue shape and correlate these movements with formant structure modulation to quantify the number of articulatory degrees of freedom and the ability to independently control these throughout a vocalization. Such studies will further elucidate the mechanisms behind seals' vocal production and dynamic vocal plasticity, ideally using quantitative modeling of vocal tract shape and comparing it with observed vocal output.

AUTHOR CONTRIBUTIONS

Conceptualization and methodology: C.R. and W.T.F.; Data curation: W.T.F., A.R., and M.G.; Formal analysis and interpretation of results: M.G., Y.J., A.R., and W.T.F.; Funding acquisition and resources: C.R., W.T.F., and A.R; Investigation: C.R. and W.T.F.; Project administration: M.G. and A.R.; Software and visualization: M.G. and Y.J.; Supervision: A.R., W.T.F., and C.R.; Validation: W.T.F., A.R., and Y.J.; Writing—original draft: M.G.; Writing—review and editing: M.G., Y.J., A.R., W.T.F., and C.R.

All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGMENTS

Ronald Schusterman invited W.T.F. to make these unique observations and recordings of harbor seal Sprouts at Long Marine Laboratory. We completed this paper some years later in honor of his hospitality and contributions to the field of animal communication. Marla Holt, Kristy Lindemann-Biolsi, Jason Mulsow, and David Kastak assisted with the collection of these data. The authors thank Ryan Jones for kindly providing the line drawings for Figure 2, Marija Spasikova for the help with data preprocessing, and Yaroslav I. Sobolev for priceless advice on data analysis.

Open access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors have no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

A video demonstrating an example of the raw data is provided in Supporting Information S1. The processed dataset used in statistical analyses is provided in Supporting Information S2.

ORCID

Maria Goncharova https://orcid.org/0000-0003-1741-224X
Yannick Jadoul https://orcid.org/0000-0003-0540-3135
Colleen Reichmuth https://orcid.org/0000-0003-0981-6842
W. Tecumseh Fitch https://orcid.org/0000-0003-1830-0928
Andrea Ravignani https://orcid.org/0000-0002-1058-0024

PEER REVIEW

The peer review history for this article is available at: https://publons.com/publon/10.1111/nyas.15189

REFERENCES

1. Fant, G. (1970). *Acoustic theory of speech production*. Walter de Gruyter.
2. Titze, I. R., & Martin, D. W. (1998). *Principles of voice production*. Acoustical Society of America.
3. Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms12739
4. Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*, *102*(2), 1213–1222. https://doi.org/10.1121/1.421048
5. Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 941–947. https://doi.org/10.1098/rspb.2004.2954
6. Fitch, W. T. (1999). Acoustic exaggeration of size in birds via tracheal elongation: Comparative and theoretical analyses. *Journal of Zoology*, *248*(1), 31–48. https://doi.org/10.1111/j.1469-7998.1999.tb01020.x
7. Reber, S. A., Janisch, J., Torregrosa, K., Darlington, J., Vliet, K. A., & Fitch, W. T. (2017). Formants provide honest acoustic cues to body size in American alligators. *Scientific Reports*, *7*(1), 1816. https://doi.org/10.1038/s41598-017-01948-1
8. Fitch, W. T., & Hauser, M. D. (2003). Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, R. R. Fay, & A. N. Popper (Eds.), *Acoustic communication*. (pp. 65–137). Springer. https://doi.org/10.1007/0-387-22762-8_3
9. Beeck, V. C., Heilmann, G., Kerscher, M., & Stoeger, A. S. (2022). Sound visualization demonstrates velopharyngeal coupling and complex spectral variability in Asian elephants. *Animals*, *12*(16), Article 16. https://doi.org/10.3390/ani12162119
10. Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1–20. https://doi.org/10.1111/j.1469-7998.2012.00920.x
11. Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology*, *280*(3), 221–236. https://doi.org/10.1111/j.1469-7998.2009.00661.x
12. Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1206–1219. https://doi.org/10.1044/1092-4388(2010/09-0083)
13. Zlegler, W., & Cramon, D. V. (1983). Vowel distortion in traumatic dysarthria: A formant study. *Phonetica*, *40*(1), 63–78. https://doi.org/10.1159/000261681
14. Roth, G., & Wake, D. (1989). Evolution of feeding in vertebrates. In D. B. Wake & G. Roth (Eds.), *Complex organismal functions: Integration and evolution in vertebrates*. (pp. 7–21). New York: Wiley-Interscience.
15. Carterette, E., Shipley, C., & Buchwald, J. (1979). Linear prediction theory of vocalization in cat and kitten. In B. Lindblom & S. Öhman (Eds.), *Frontiers in speech communication research* (pp. 245–257).
16. Fitch, W. T. (2000). The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica*, *57*(2–4), 205–218. https://doi.org/10.1159/000028474
17. Hauser, M. D., Evans, C. S., & Marler, P. (1993). The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. *Animal Behaviour*, *45*(3), 423–433. https://doi.org/10.1006/anbe.1993.1054
18. Riede, T., Bronson, E., Hatzikirou, H., & Zuberbühler, K. (2005). Vocal production mechanisms in a non-human primate: Morphological data and a model. *Journal of Human Evolution*, *48*(1), 85–96. https://doi.org/10.1016/j.jhevol.2004.10.002
19. Shipley, C., Carterette, E. C., & Buchwald, J. S. (1991). The effects of articulation on the acoustical structure of feline vocalizations. *Journal of the Acoustical Society of America*, *89*(2), 902–909. https://doi.org/10.1121/1.1894652
20. Elemans, C. P. H., Jiang, W., Jensen, M. H., Pichler, H., Mussman, B. R., Nattestad, J., Wahlberg, M., Zheng, X., Xue, Q., & Fitch, W. T. (2024). Evolutionary novelties underlie sound production in baleen whales. *Nature*, *627*(8002), 123–129. https://doi.org/10.1038/s41586-024-07080-1
21. Madsen, P. T., Siebert, U., & Elemans, C. P. H. (2023). Toothed whales use distinct vocal registers for echolocation and communication. *Science*, *379*(6635), 928–933. https://doi.org/10.1126/science.adc9570
22. Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1477), 1669–1675. https://doi.org/10.1098/rspb.2001.1704
23. Frey, R., Volodin, I., Volodina, E., Soldatova, N. V., & Juldaschev, E. T. (2011). Descended and mobile larynx, vocal tract elongation and rutting roars in male goitred gazelles (Gazella subgutturosa Güldenstaedt, 1780). *Journal of Anatomy*, *218*(5), 566–585. https://doi.org/10.1111/j.1469-7580.2011.01361.x
24. Frey, R., Volodin, I., Volodina, E., Carranza, J., & Torres-Porras, J. (2012). Vocal anatomy, tongue protrusion behaviour and the acoustics of rutting roars in free-ranging Iberian red deer stags (*Cervus elaphus* hispanicus). *Journal of Anatomy*, *220*(3), 271–292. https://doi.org/10.1111/j.1469-7580.2011.01467.x

25. Frey, R., Volodin, I. A., Volodina, E. V., Efremova, K. O., Menges, V., Portas, R., Melzheimer, J., Fritsch, G., Gerlach, C., & von Dörnberg, K. (2020). Savannah roars: The vocal anatomy and the impressive rutting calls of male impala (*Aepyceros melampus*)—Highlighting the acoustic correlates of a mobile larynx. *Journal of Anatomy*, *236*(3), 398–424. https://doi.org/10.1111/joa.13114
26. Riede, T. (2011). Subglottal pressure, tracheal airflow, and intrinsic laryngeal muscle activity during rat ultrasound vocalization. *Journal of Neurophysiology*, *106*(5), 2580–2592. https://doi.org/10.1152/jn.00478.2011
27. Volodin, I., Volodina, E., Frey, R., Carranza, J., & Torres-Porras, J. (2013). Spectrographic analysis points to source–filter coupling in rutting roars of Iberian red deer. *Acta Ethologica*, *16*(1), 57–63. https://doi.org/10.1007/s10211-012-0133-1
28. Hauser, M. D., & Schön Ybarra, M. S. (1994). The role of lip configuration in monkey vocalizations: Experiments using xylocaine as a nerve block. *Brain and Language*, *46*(2), 232–244. https://doi.org/10.1006/brln.1994.1014
29. Titze, I. (1994). Fluctuations and perturbations in vocal output. In I. Titze (Ed.), *Principles of voice production* (pp. 209–306).
30. Fahlman, A., Moore, M. J., & Garcia-Parraga, D. (2017). Respiratory function and mechanics in pinnipeds and cetaceans. *Journal of Experimental Biology*, *220*(10), 1761–1773.
31. Parisi, I., De Vincenzi, G., Torri, M., Papale, E., Mazzola, S., Bonanno, A., & Buscaino, G. (2017). Underwater vocal complexity of Arctic seal *Erignathus barbatus* in Kongsfjorden (Svalbard). *Journal of the Acoustical Society of America*, *142*(5), 3104–3115. https://doi.org/10.1121/1.5010887
32. Reichmuth, C., & Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology*, *28*, 66–71. https://doi.org/10.1016/j.conb.2014.06.011
33. Stansbury, A. L., & Janik, V. M. (2019). Formant modification through vocal production learning in gray seals. *Current Biology*, *29*(13), 2244–2249.
34. Schusterman, R. J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. In D. K. Oller, & U. Griebel (Eds.), *Evolution of communicative flexibility: Complexity, creativity, and adaptability in human and animal communication*. (pp. 41–70). MIT Press. https://doi.org/10.7551/mitpress/9780262151214.003.0003
35. Shapiro, A. D., Slater, P. J. B., & Janik, V. M. (2004). Call usage learning in gray seals (*Halichoerus grypus*). *Journal of Comparative Psychology*, *118*(4), 447–454. https://doi.org/10.1037/0735-7036.118.4.447
36. Michaelis, H., & Jones, D. (2003). *A phonetic dictionary of the English language* (Vol. 2). Psychology Press.
37. Casey, C., Sills, J. M., Knaub, S., Sotolotto, K., & Reichmuth, C. (2021). Lifelong patterns of sound production in two seals. *Aquatic Mammals*, *47*(5), 499–514.
38. Reichmuth, C. (2023). The life semi-aquatic: Harbor seal sprouts and milestones in marine bioacoustics. *Journal of the Acoustical Society of America*, *153*, (3_supplement), A308. https://doi.org/10.1121/10.0018953
39. Sills, J. M., Southall, B. L., & Reichmuth, C. (2014). Amphibious hearing in spotted seals (*Phoca largha*): Underwater audiograms, aerial audiograms and critical ratio measurements. *Journal of Experimental Biology*, *217*(5), 726–734. https://doi.org/10.1242/jeb.097469
40. Hanggi, E. B., & Schusterman, R. J. (1994). Underwater acoustic displays and individual variation in male harbour seals, *Phoca vitulina*. *Animal Behaviour*, *48*(6), 1275–1283.
41. Nowak, L. J. (2021). Observations on mechanisms and phenomena underlying underwater and surface vocalisations of grey seals. *Bioacoustics*, *30*(6), 696–715. https://doi.org/10.1080/09524622.2020.1851298
42. Buschmeier, H., & Wlodarczak, M. (2013). TextGridTools: A TextGrid processing and analysis toolkit for Python. *Tagungsband Der 24. Konferenz Zur Elektronischen Sprachsignalverarbeitung (ESSV 2013)*.
43. Boersma, P., & Weenik, D. (2022). *Praat: Doing phonetics by computer [Computer program]*. (6.2.06) [Computer software]. https://www.praat.org
44. Jadoul, Y., Thompson, B., & Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001
45. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), Article 7825. https://doi.org/10.1038/s41586-020-2649-2
46. van Kemenade, H., Murray, A., Clark, J. A., Karpinsky, A., Baranovič, O., Gohlke, C., Dufresne, J., Schmidt, D., Kopachev, K., Houghton, A., Mani, S., Landey, S., Ware, J., Piolie, Douglas, J., … Base, M. (2022). *python-pillow/Pillow: 9.2.0*. [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6788304
47. Silvester, S., Tanbakuchi, A., Müller, P., Nunez-Iglesias, J., Harfouche, M., McCormick, M., Ladegaard, A., Rai, A., Smith, T. D., Konowalczyk, M., Lee, A., Klein, A., Nises, J., Vaillant, G. A., Barnes, C., Zulko, … Dusold, C. (2020). *Imageio/imageio v2.8.0*. [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.3674133
48. Fry, D. B. (1979). *The physics of speech*. Cambridge University Press.
49. Vorperian, H. K., Kurtzweil, S. L., Fourakis, M., Kent, R. D., Tillman, K. K., & Austin, D. (2015). Effect of body position on vocal tract acoustics: Acoustic pharyngometry and vowel formants. *Journal of the Acoustical Society of America*, *138*(2), 833–845. https://doi.org/10.1121/1.4926563
50. Lindblom, B. E. F., & Sundberg, J. E. F. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America*, *50*(4B), 1166–1179. https://doi.org/10.1121/1.1912750
51. Litman, R. S., Wake, N., Chan, L.-M. L., McDonough, J. M., Sin, S., Mahboubi, S., & Arens, R. (2005). Effect of lateral positioning on upper airway size and morphology in sedated children. *Anesthesiology*, *103*(3), 484–488. https://doi.org/10.1097/00000542-200509000-00009
52. Ono, T., Otsuka, R., Kuroda, T., Honda, E., & Sasaki, T. (2000). Effects of head and body position on two- and three-dimensional configurations of the upper airway. *Journal of Dental Research*, *79*(11), 1879–1884. https://doi.org/10.1177/00220345000790111101
53. Pae, E.-K., Lowe, A. A., Sasaki, K., Price, C., Tsuchiya, M., & Fleetham, J. A. (1994). A cephalometric and electromyographic study of upper airway structures in the upright and supine positions. *American Journal of Orthodontics and Dentofacial Orthopedics*, *106*(1), 52–59. https://doi.org/10.1016/S0889-5406(94)70021-4
54. Pevernagie, D. A., Stanson, A. W., Sheedy, P. F., Daniels, B. K., & Shepard, J. W. (1995). Effects of body position on the upper airway of patients with obstructive sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, *152*(1), 179–185. https://doi.org/10.1164/ajrccm.152.1.7599821
55. Shiller, D. M., Ostry, D. J., & Gribble, P. L. (1999). Effects of gravitational load on jaw movements in speech. *Journal of Neuroscience*, *19*(20), 9073–9080. https://doi.org/10.1523/JNEUROSCI.19-20-09073.1999
56. Van Holsbeke, C. S., Verhulst, S. L., Vos, W. G., De Backer, J. W., Vinchurkar, S. C., Verdonck, P. R., van Doorn, J. W. D., Nadjmi, N., & De Backer, W. A. (2014). Change in upper airway geometry between upright and supine position during tidal nasal breathing. *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, *27*(1), 51–57. https://doi.org/10.1089/jamp.2012.1010
57. Schneider, R., & Kükenthal, W. G. (1964). *Der larynx der säugetiere*. de Gruyter.
58. Lee, J., Shaiman, S., & Weismer, G. (2016). Relationship between tongue positions and formant frequencies in female speakers. *Journal*

of the Acoustical Society of America, *139*(1), 426–440. https://doi.org/10.1121/1.4939894

59. Fernandez, A. A., Burchardt, L. S., Nagy, M., & Knörnschild, M. (2021). Babbling in a vocal learning bat resembles human infant babbling. *Science*, *373*(6557), 923–926. https://doi.org/10.1126/science.abf9279

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Goncharova, M., Jadoul, Y., Reichmuth, C., Fitch, W. T., & Ravignani, A. (2024). Vocal tract dynamics shape the formant structure of conditioned vocalizations in a harbor seal. *Ann NY Acad Sci.*, 1–10. https://doi.org/10.1111/nyas.15189