UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Why do wolf howls sound like "uw"? Behavioral observations and exploratory modeling

Permalink

https://escholarship.org/uc/item/2sk1z9nk

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Ekström, Axel
Amici, Federica
Friedrichs, Daniel
et al.

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Why do wolf howls sound like "uw"? Behavioral observations and exploratory modeling

Axel Ekström (axel.ekstrom@su.se)
Centre for Cultural Evolution, Department of Psychology, Stockholm University, Stockholm, Sweden
Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

Federica Amici
Institute of Biology, University of Leipzig, Leipzig, Germany

Daniel Friedrichs
Department of Computational Linguistics, University of Zurich, CH–8050 Zurich, Switzerland

Linda Oña
Institute of Biology, University of Leipzig, Leipzig, Germany

Abstract

We report observations of Hudson Bay wolves (*Canis lupus hudsonicus*) producing howl–like calls accompanied by visible dynamic tongue movements. We show that the wolves actively changed the shape of their vocal tract through apical tongue gestures (i.e., the tongue tip arches anteriorly toward the palate), a phenomenon that changes vowel quality in human speech. A combination of video content analysis and articulatory modeling indicates that the apical lingual gesture changes vocal-tract resonances along the close–open dimension, producing a consistent perceptual shift from a schwa–like [ə] quality toward a close central rounded quality [ʉ]. These findings provide evidence that a non-primate animal can perform dynamic lingual filtering during vocal production.

Keywords: iconicity; speech; animal vocalization

Introduction

Vocal communication depends not only on sound generation in the larynx but also on how the supralaryngeal vocal tract shapes that sound. In humans, this is most clearly illustrated in vowel production: the voice "source" from the vocal folds—measured as fundamental frequency (f_0) – is "filtered" by the changing shape of the supralaryngeal vocal tract. By imposing narrow constrictions on airflow through the various articulators — the tongue, lips, teeth, hard and soft palates, uvula, pharyngeal wall, and glottis — human speakers attenuate the spectral energy peaks (termed formants) in the acoustic signal in largely predictable ways (Fant, 1971; Stevens, 2000). In traditional phonetics nomenclature, the first formant (F1) is held to correspond to the closeness of constriction, or mouth closure; the second formant (F2) to the front-to-back tongue position; and the third formant (F3) to the degree of lip rounding. For example, when articulating close front unrounded vowel [i] (the vowel in "see"), the lingual dorsum is typically "close" to the palate (shifting down F1) in the "front" of the oral cavity (shifting up F2).

In many mammalian vocalizations, formant structure is present but is often shaped mainly by changes in effective vocal-tract length (e.g., jaw opening, lip configuration, laryngeal descent) rather than by precise, speech-like control of tongue posture. As a result, many calls are produced with a comparatively neutral supralaryngeal configuration, yielding a centralized, [ə]–like ("schwa–like") vowel quality. This is one reason why a single "neutral" vocal-tract setting can approximate a wide range of mammalian calls in synthesis frameworks (Moore, 2016).

To date, few known exceptions to the "schwa" pattern in animal vocalizations, where disparate vowel-like qualities are achieved through visible maneuvers of the articulators. Non-schwa-like vowel-like calls have been reported for non-human primates, but the exact configurations employed by primates to achieve variable "vowel-like" properties are uncertain. Schötz (2020) has noted vowel-like sounds in domestic cats (*Felis catus sylvestris*), although these determinations were based on correlations between jaw heights and auditory impressions, rather than formant estimates, and thus do not necessarily imply the use of the tongue (Shipley, Carterette, & Buchwald, 1991; Ekström, Vila, Schötz, & Edlund, 2024). Other data show changes to formants, from involuntary tongue shape distortions to differences in posture reflecting gravitational pulls (Goncharova, Jadoul, Reichmuth, Fitch, & Ravignani, 2024).

From the point of view of articulatory phonetics, even though animals such as felines (Shipley et al., 1991; Schötz, 2020; Ekström et al., 2024), cervids (McElligott, Birrer, & Vannoni, 2006) and pinnipeds (Goncharova et al., 2024) may produce variable formant patterns, the production of these dispersions typically reflects shortening (through low jaw heights or high larynx positions) or elongation (through high jaw or low larynx positions) of the vocal tract (B. E. Lindblom & Sundberg, 1971). Therefore, they do not diverge from the neutral "schwa" configuration. In speech, distinctive vowel qualities reflect voluntary and replicable tongue shapes (not formant dispersions *per se*, which can vary widely between speakers). Here, we investigate articulatory mechanisms underlying wolf (*Canis lupus hudsonicus*) howls, and report the first evidence of dynamic tongue movements affecting vowel–like quality in a carnivoran.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Methods

Data collection

We collected our data at Zoo Osnabrück (Germany) during two periods: October 2017 (by LO) and August-November 2023 (by MD and HB). The wolf enclosure entailed a surface of 4,500 m^2, including an indoor enclosure of approximately 30 m^2. All individuals in the group were born in captivity. During the initial data collection, the group comprised six wolves (2 females and 4 males) aged between 0.5 and 7 years. At the onset of data collection, a new pack was established. Four wolves from Amsterdam Zoo were integrated with two females from Zoo Olomouc at Osnabrück Zoo. By the second recording session, the group had expanded to twelve individuals (6 males and 6 females) following the birth of two litters in 2019 and 2020, and the passing of three original members.

Howling events were recorded opportunistically during daytime and mainly triggered by external sounds such as ambulances, trucks, and alarms outside the zoo or announcements delivered through loudspeakers (Theberge & Theberge, 2022). From a comprehensive dataset of approximately 570 hours of video footage, we identified and segmented approximately 51 minutes containing howling behaviors. Video was recorded using a Panasonic HC-V785 camera at a resolution of 1920×1080 pixels and 50 frames per second. All videos were taken from a platform overlooking the enclosure; as the individuals moved around, the distance to the animals varied between approximately 5 and 30 meters. The audio was recorded in stereo format with a bit rate of 125 kbps and a sample rate of 48 kHz using the camera's built-in stereo microphone with wind protection.

Acoustic and phonetic assessment

Wolf howls often have high f_0, and our recordings were made with a camera microphone using compressed audio (125 kbps). For these reasons, we did not attempt parametric formant tracking from the recordings. Instead, we used a conservative procedure focused on (i) selecting tokens suitable for perceptual labeling, and (ii) qualitative spectrographic inspection. A subset of howls was retained for phonetic labeling only when (a) a single caller dominated the recording, (b) overlapping chorusing was minimal, and (c) the tongue gesture could be time-aligned with the vocalization from the video. For these tokens, a trained phonetician provided an impressionistic IPA-based label of the dominant vowel-like quality while inspecting spectrograms, and annotated the time span corresponding to the sustained apical gesture. We treat these labels as impressionistic (single-coder) and use them as converging evidence alongside video observations and articulatory modeling.

Video content analysis

Simultaneous analysis of video and audio material allows for the analysis of articulatory configurations plausibly associated with the apparent vowel-like quality observed in howling. The purpose of the video content analysis procedure was to identify and categorize lingual gestures associated with apparent vowel-like qualities in howling behavior. Videos and vowel-like qualities were annotated simultaneously by the same person, who used the ELAN software (Sloetjes & Wittenburg, 2008) to manually annotate each instance where the tongue was visible and visibly moved during the production of the call. Instances where only brief apical "flickering" (i.e., momentary tongue arching) were excluded.

Vocal tract models

Vocal tract data Vocal tract length (VTL) is a primary determinant of formant frequencies. While VTL data has been reported for various domestic dog (*C. l. familiaris*) breeds (Riede & Fitch, 1999), there is, to our knowledge no equivalent data specific to wolves. Based on the strong positive correlation between the skull length (SL) and VTL (r = 0.962, $p < .001$) reported for domestic dogs (Riede & Fitch, 1999), we estimated the VTL for Hudson Bay wolves using available measurements of canine craniofacial features. According to van Valkenburgh et al. (2014), the SL of a *C. l. lupus* specimen is 23.6 cm – which falls within the range reported for German shepherds, suggesting continuity in relevant morphological parameters. Based on averaged data from three German shepherds (Riede & Fitch, 1999, p. 2862), we derived the following scaling factor:

$$\frac{VTL}{SL} = 0.9 \tag{1}$$

Applying this formula to the SL of an arctic wolf specimen reported by van Valkenburgh and colleagues, we estimated:

$$VTL = 23.6 \cdot 0.9 = 21.24 \tag{2}$$

Note that this procedure likely underestimates the wolf VTL. Riede and Fitch (1999, p. 2861) measured SL as "the distance between the front of the incisors and the protuberantia occipitalis externa of the os occipitale", while van Valkenburg and colleagues (2014, p. 2070) defined it as "the distance between the anterior alveolar margin of the central incisors and the posterior edge of the occipital condyle." The latter measurement, derived from more widely spaced anatomical landmarks, produces a comparatively deflated value. Judging by the position of the *protuberantia occipitalis externa* in the illustration provided by van Valkenburgh et al. (2014, p. 2067), the definition used by Riede and Fitch would produce an SL approximately 1-2 cm longer. However, the arctic wolf, from which our SL measurement is derived, is a larger-bodied taxon than the Hudson Bay wolf, which may partially offset the underestimation.

Computer model We employed articulatory–acoustic modeling to to evaluate whether the observed apical tongue gesture could plausibly shift vocal-tract resonances in the

direction suggested by vowel perception. We used the *TubeN* software (Zhang et al., 2024) which implements the Liljencrants–Fant tube-model algorithm for simulating vocal-tract area transfer functions (1975).

Disregarding losses, the computational representation of one segment is as follows:

$$\mathrm{a_n} = Z_\mathrm{n} jtg\ (\tfrac{1}{2} \cdot \frac{wl_n}{c}) \tag{3}$$

$$\mathrm{b_n} = \frac{Z_\mathrm{n}}{jsin(wl_n/c)} \tag{4}$$

$$\mathrm{a_n} + \mathrm{b_n} = \frac{Z_\mathrm{n}}{jtg(wl_n/c)} \tag{5}$$

$$\mathrm{Z_n} = \rho\ c/A_n \tag{6}$$

where l_n corresponds to the length, and A_n to cross-sectional area, of the N_{th} segment. Radiation inductance is computed as:

$$b_o = \frac{jwl_o}{A_o}, l_o = 0.8\sqrt{\frac{A_o}{\pi}} \tag{7}$$

The program implements an internal-end correction procedure:

$$k_n = \frac{A(n)}{A(n-1)} B_n = k_n \cdot \frac{Sn}{Sn-1} \tag{8}$$

which is repeated for each segment, until all segments are processed. The final value Y is computed as

$$Y = C_n + B_n \cdot C_{n-1}. \tag{9}$$

The Liljencrants-Fant code computes the resonance frequencies of the vocal tract using a transfer determinant method applied to a model of the tract as a series of connected acoustic tubes. For a given test frequency, it recursively evaluates how pressure waves propagate through each tube segment, taking into account their length L and cross-sectional areas A. At each recursive step, the program produces an updated system determinant (y) that reflects how energy builds or cancels across the sequence. As such, the software models a closed-to-open tube or sequence of tubes, allowing for the simulation of various articulatory configurations, and predicts formants for each configuration. We here opted for a conservative tube modeling strategy, reflecting the rarity of previous applications to animal calls (Shipley et al., 1991) More exhaustive efforts to model vowel-like vocal production in non-human animals are necessary to determine appropriate considerations.

Shipley et al. (1991, p. 906) write, "One way that jaw movement could affect harmonic structure is in shortening the length of the vocal tract, thereby raising its characteristic resonances. As the jaw lowers, the lips move apart at the side of the mouth and the effective length of the tract is reduced". That is, the acoustic effects of jaw lowering and raising may, in theory, be equivalent to shortening (low jaw) and elongation (high jaw) of the total VTL. The assumption, as a minimum, is thus that jaw position is inversely related to effective vocal tract length. Shipley et al.(Shipley et al., 1991) modeled the jaw movements of a cat as equivalent of to the "flare" of a Bessel horn; however, in our case, this was not appropriate, specifically because the tongue gestures were likely to offset the transfer function of the would-be horn. To our knowledge there exists no attempt to model any kind of horn flaring with a flare-internal obstruction of the kind we observed. Nonetheless, the literature on the acoustics of musical instruments may be useful in future attempts (Bängtsson, Noreland, & Berggren, 2003).

The base vocal tract was modeled as closed-to-open tube with a total length of 21.24 cm, corresponding to our VTL estimate. Based on visual observations, we placed the constriction between 8 cm and 10 cm from the "lips" end, corresponding to the position near the premolars as per van Valkenburgh et al.(van Valkenburgh et al., 2014) The constriction length was varied from 0.5 cm to 2 cm in increments of 0.1 cm, and the degree of constriction was varied from 0.2 cm^2 (minor constriction) to 3.8 cm^2 (significant constriction) in increments of 0.1 cm^2. To maintain a constant VTL of 21.24 cm, the length of the remaining "lips" section was inversely adjusted with constriction length. Finally, for "abrupt" transitions constrictions it was necessary to include an additional end correction, derived from Ingård (1953)

$$\delta_i \simeq 0.48(A)^{\frac{1}{2}}(1 - 1.25\xi) \tag{10}$$

which was, following Fant (1971, p. 36), applied where $A_{n+1} < A_n \cdot 0.15$. That is, where the area A of a preceding segment n was 15% or less of the area of the following segment, the length of the segment was adjusted.

The resultant tube model consisted of four sections, with sections dorsal and anterior to the constrictions having a constant cross-sectional area of 4 cm^2. In addition, lowering the jaw will typically narrow the pharyngeal cavity. To approximate this, the final segment was modeled as 75% of the general. This issue bears additional consideration in the future, given its ubiquity in speech. Finally, for vowel-like calls, the acoustic effect of an additional resonator (i.e., nasal cavity) would result in pronounced spectral valleys (or anti-resonances, anti-formants). While animals likely do not close off the velum as readily as human speakers (Lieberman, 2006), there are indications that they may do so through a temporary re-arrangement of vocal anatomy during loud

calls, such as howling (Fitch, 2010). Accordingly, we modeled vocal tract area transfer functions as non-nasalized filtering.

A second round of modeling was also performed, maintaining identical constriction settings, except for implementing a notch at the "lips" end of the tube section, representing a lowered jaw position. Effects of notching (or "lip spreading") can be emulated by implementing a shorter uniform segment in place of the notch (B. Lindblom, Sundberg, Branderud, & Djamshidpey, 2007). To our knowledge, no work has described the extent of jaw lowering available to canines, nor the extent actively employed in howling. We implemented a 5 cm notch as a 1.75 cm uniform segment after Lindblom et al. (2007), resulting in an effective VTL of 17.99 cm; all other settings were kept identical. Our "notch" is likely an underestimate; however, lacking both buccolabial tract measurements from live wolves and an empirical acoustic basis for inferring the effects of notching past 5 cm – the deepest notched tested by Lindblom and colleagues – we opted for settings that were conservative but replicable and motivated. As such, this second set of models was highly programmatic and served to demonstrate the consistency of articulatory-acoustic relationships under scrutiny for different jaw positions. Contrasts observed would likely become more realistic and perceptually distinctive when informed by additional anatomical and acoustic data.

Results

Video content analysis

We observed multiple instances (N = 31) of apical tongue movements associated with howling behavior in Hudson Bay wolves. At initiation of each howl, the tongue was visibly flat against the mandibular floor, and positioned within the mandibular dental arch. As the howl progressed, the tongue body retracted, and the apex arched dorsoanteriorly, elevating away from the mandibular canines. This apical posture was typically sustained throughout the duration of the howl, as evidenced by the visible lingual frenulum. Vowel-like quality could be reliably determined in 9 utterances, where the individual was sufficiently close to the microphone and minimal overlapping vocalizations were present. In these cases, vowel–like quality consistently resembled the close central rounded vowel [ʉ]. This vowel is uncommon in English dialects (though may occur in e.g., Australian and New Zealand English) but commonly appears in languages such as Swedish (e.g., *duk* [dʉːk], tablecloth). In traditional phonetics nomenclature, [ʉ] employs a "close" tongue position that represents a midpoint between close back rounded vowel [u] (where stricture is in the back of the oral cavity) and close front rounded vowel [y] (where stricture is in the front). It typically exhibits a distinct formant pattern where F1 $\approx$ 300-400 Hz, and F2 $\approx$ 1200-1400 Hz (Riad, 1997; Asu, Schötz, & Kügler, 2009).

Vocal tract model

We modeled wolf apical constriction as a series of tube segments, where constriction was intended to simulate the effect of apical tongue movements on formants. We also performed a second round of modeling, based on the dictum that opening the jaw constitutes a "notch" in the anterior-most segment, which is functionally equivalent to a shorter segment than its un-notched counterpart (B. Lindblom et al., 2007). Effectively, our models tested the effect of constriction using two models, with different anteriormost segment lengths (i.e., varying the distance between the constriction itself and the termination of the tube segment at the "lip" end).

Our articulatory models revealed a nonmonotonic relationship between constriction area and formant frequencies. Specifically, greater stricture in the vocal tract facilitated an increasingly drastic downward shift in F1. F2 exhibited a similar trend, although the magnitude of change was smaller; for notched tubes, modeling a shorter effective vocal tract, the effect was minute, and its high frequency implies that the contribution of the change to vowel quality is minor, compared to the contribution of changes in F1 (Fig. 1). In addition, we observed a nonmonotonic articulatory-acoustic relationship between constriction area and F1, F3, and (for the uniform models) F2, starting at approximately 1.5 cm^2. Further reductions in constriction area resulted in progressively larger shifts in formant frequencies (Stevens, 2000). The greater shift in F3 compared to F2, resulted in convergence between F2 and F3. The implemented stricture shifted the perceived vowel quality from a mid central vowel [ə] (or "schwa") to close central rounded vowel [ʉ], consistent with our impressionistic phonetic labels. Notched tube models, which incorporated a 5-cm jaw lowering simulation, produced a more significant contrast in vowel-like quality between non-constricted and constricted forms compared to uniform tube models. However, varying the constriction length did not result in similar displacements in formant frequencies.

Discussion

To our knowledge, this constitutes the first report of dynamic lingual vocal tract "filtering" in a carnivoran. In our sampled howls, the apparent vowel-like quality (for relatively low-f_0 calls) was consistent with perception of [ʉ] not [u], reinforcing the interpretation of a more "central" constriction. While we did not inspect brief stricture events (where apical contact or closeness to the palate was brief in time and followed by rapid displacements in apical position), phonetic results of these movements are likely akin to an /ɹ/-colored vowel [ɚ], with a likely corresponding downward shift of F3 (Ladefoged & Maddieson, 1996). Such "flickering" may be responsible for the "raow"-like quality sometimes perceived in howling (pers. obs.); this is also consistent with our articulatory-acoustic models, which predicted a marked decrease in F3 with increased stricture. In events of prolonged apical closeness, vocalizations may also take on a quality similar to the voiced postalveolar approximant [ɹ] (the first consonant in

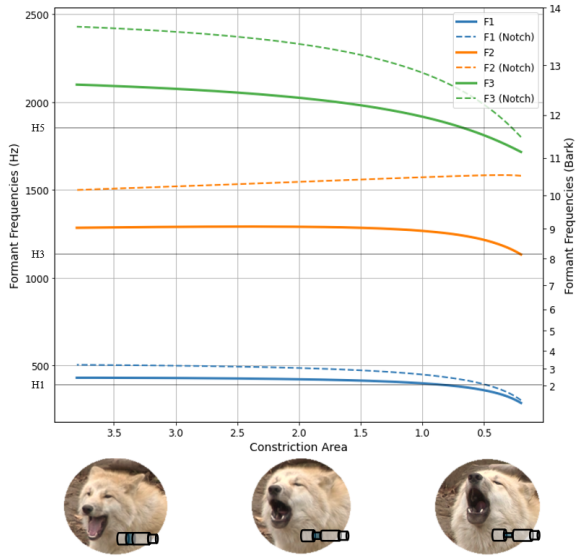


Figure 1: The effect of constriction area on F1 and F3 is non-monotonic, such that decreasing the area of constriction induces progressively greater shifts in frequency. In comparison, constriction length is of relatively minor consequence.

"red"). Results of both auditory and video-based procedures are thus consistent with a central tongue position. In our models, the stricture shifted down F1 and F3 more than F2, facilitating a quality similar to close central rounded vowel [ʉ]. Ours is the first evidence of a tongue movement facilitating a change to vowel-like quality; however, this movement is not executed within the same range of constraints as human speech production. Works aiming to compare vocal production across species, should seek to specify such distinct constraints on articulation. Our findings illustrate that the comparison of even ostensibly "vowel-like" phenomena demands explicit consideration of species-unique constraints on production: apical gestures observed in our wolf subjects are not identical to those observed in human speakers producing the "same" vowel. Given the relative extremity of the gestures employed and the apparent social context in which they are used, it is possible to speculate on causal mechanisms.

It is possible that wolves, by retracting the larynx into the lower throat (rather than simply narrowing the pharynx), facilitate the robustness of the spectral characteristics of the call by removing spectral valleys imposed by the nasal tract (Havel, Sundberg, Traser, Burdumy, & Echternach, 2023). Howls are widely believed part of safeguarding a pack's territory (Harrington & Mech, 1979; Zaccaroni et al., 2012; Mazzini, Townsend, Virányi, & Range, 2013). It is possible that a variety of source-filter interactions, unaccounted for in our work, are likely relevant to future analysis and to more in-depth understanding of group dynamics in howling. Howling is widely believed to contribute to territorial defense (Harrington & Mech, 1979; Zaccaroni et al., 2012; Mazzini et al., 2013), and a richer set of source–tract interactions than we model here (including possible interactions between jaw opening, supralaryngeal constriction, and the harmonic structure imposed by high f_0) may therefore be relevant for understanding both the acoustics of the call and the dynamics of group howling. For example, in speech, lowering the jaw contributes to changes in F1 (Fant, 1971), and adds to impressions of prominence (Erickson, 2002). Articulatory configurations employed in howling may transmit an "honest signal" with information about the qualities of the pack. Finally, while our dataset includes only one of over 30 extant grey wolf subspecies, we are not aware of any anatomical or behavioral reason that the apical articulatory gestures observed should be unique to the Hudson Bay subspecies. Future work may confirm the phenomenon across the genus.

Our work has several limitations which may be addressed in future research. First, howling is rare, occurring in $< 0.2\%$ of our video corpus of over 570 hours. In addition, our analysis required the howling individual to be facing the camera at the moment of recording, and while we collected a substantial audio data sample of howling, the wolves' articulators were rarely outwardly visible. The rarity of the behavior and constraints of our methodology effectively deflated our usable sample size. Second, simulations reported here are rough simulacra, illustrating a principled relationship. Our work shows that perturbing a uniform tube of approximately the length of the gray wolf vocal tract at a position inferred from observed apical constrictions, results in a closer acoustic approximation of close central rounded vowel [ʉ]. These simulations are not, however, to be taken as anatomically accurate.

Our articulatory models necessarily rely on anatomical assumptions. In particular, we extrapolated aspects of skull–vocal-tract relationships from domestic dogs. While domestic dogs and grey wolves are categorized as the same species, dogs have been subject to extensive selective breeding, which is cause for caution. Our placements of apical constriction were based on a combination of in-vivo observations (of the lingual apex approaching the palate at approximately the position of the maxillary molars) and measurements of skull anatomical features provided for an arctic wolf (van Valkenburgh et al., 2014, p. 2069), a slightly larger subspecies than Hudson Bay wolves. Our findings may be verified through collaborations with, for example, wildlife reserves or national parks, where wolves' vocalization behavior can be sampled in their natural habitats (Theberge & Theberge, 2022), and anatomical data can be obtained opportunistically from deceased individuals under appropriate permits and ethical oversight. Were such anatomical data to be provided, more elaborate articulatory modeling may be conducted to confirm our findings.

Acknowledgments

AE was funded through the Swedish Research Council (2025–00209_VR).

References

Asu, E. L., Schötz, S., & Kügler, F. (2009). The acoustics of estonian swedish long close vowels as compared to central swedish and finland swedish. In *Fonetik 2009: The xxiith swedish phonetics conference* (pp. 54–59). Department of Linguistics, Stockholm University.

Bängtsson, E., Noreland, D., & Berggren, M. (2003). Shape optimization of an acoustic horn. *Computer Methods in Applied Mechanics and Engineering*, *192*(11-12), 1533–1571. doi: 10.1016/S0045-7825(02)00656-4

Ekström, A. G., Vila, L. C., Schötz, S., & Edlund, J. (2024). A single formant explicates the ubiquity of "meow". In *Proceedings of the 4th international workshop on vocal interactivity in-and-between humans, animals and robots (vihar 2024)* (pp. 74–78).

Erickson, D. (2002). Articulation of extreme formant patterns for emphasized vowels. *Phonetica*, *59*(2-3), 134–149. doi: 10.1159/000066067

Fant, G. (1971). *The acoustic theory of speech production: With calculations based on x-ray studies of russian articulations*. The Hague: Mouton.

Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.

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. *nnals of the New York Academy of Sciences*. doi: https://doi.org/10.1111/nyas.15189

Harrington, F. H., & Mech, L. D. (1979). Wolf howling and its role in territory maintenance. *Behaviour*, *68*, 207-249. doi: https://doi.org/10.1163/156853979X00322

Havel, M., Sundberg, J., Traser, L., Burdumy, M., & Echternach, M. (2023). Effects of nasalization on vocal tract response curve. *Journal of Voice*, *37*, 339–347. doi: https://doi.org/10.1016/j.jvoice.2021.02.013

Ingard, U. (1953). On the theory and design of acoustic resonators. *The Journal of the Acoustical Society of America*, *25*(6), 1037–1061. doi: 10.1121/1.1907235

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell Publishers.

Lieberman, P. (2006). *Towards an evolutionary biology of language*. Harvard University Press.

Liljencrants, J., & Fant, G. (1975). Computer program for VT-resonance frequency calculations. *STL-QPSR*, *16*, 15–21.

Lindblom, B., Sundberg, J., Branderud, P., & Djamshidpey, H. (2007). On the acoustics of spread lips. In *In proceedings of fonetik 2007* (pp. 13–16). Stockholm, Sweden: TMH-QPSR, 50.

Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *4B*(50), 1166-1179. doi: https://doi.org/10.1121/1.1912750

Mazzini, F., Townsend, S. W., Virányi, Z., & Range, F. (2013). Wolf howling is mediated by relationship quality rather than underlying emotional stress. *Current Biology*, *23*. doi: https://doi.org/10.1016/j.cub.2013.06.066

McElligott, A. G., Birrer, M., & Vannoni, E. (2006). Retraction of the mobile descended larynx during groaning enables fallow bucks (Dama dama) to lower their formant frequencies. *Journal of Zoology*, *270*, 340–345. doi: https://doi.org/10.1111/j.1469-7998.2006.00144.x

Moore, R. K. (2016). A real-time parametric general-purpose mammalian vocal synthesiser. In *Interspeech* (p. 2636–2640). International Speech Communication Association.

Riad, T. (1997). *Svensk fonologikompendium*. University of Stockholm.

Riede, T., & Fitch, W. T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (Canis familiaris). *Journal of Experimental Biology*, *202*, 2859–2867. doi: https://doi.org/10.1242/jeb.202.20.2859

Schötz, S. (2020). Phonetic variation in cat–human communication. In *Pets as sentinels, forecasters and promoters of human health* (pp. 319–347). Springer. doi: https://doi.org/10.1007/978-3-030-30734-$9_1$4

Shipley, C., Carterette, E. C., & Buchwald, J. S. (1991). The effects of articulation on the acoustical structure of feline vocalizations. *The Journal of the Acoustical Society of America*, *2*(89), 902–909. doi: https://doi.org/10.1121/1.1894652

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category - ELAN and ISO DCR. In *Proceedings of the 6th international conference on language resources and evaluation (lrec 2008)*.

Stevens, K. N. (2000). *Acoustic phonetics*. MIT Press.

Theberge, J. B., & Theberge, M. T. (2022). Triggers and consequences of wolf (Canis lupus) howling in yellowstone national park and connection to communication theory. *Canadian Journal of Zoology*, *100*, 799-809. doi: https://doi.org/10.1139/cjz-2022-0043

van Valkenburgh, B., Pang, B., Bird, D., Curtis, A., Yee, K. K., Wysocki, C. J., & Craven, B. A. (2014). Respiratory and olfactory turbinals in feliform and caniform carnivorans: The influence of snout length. *The Anatomical Record*, *297*, 2065–2079. doi: https://doi.org/10.1002/ar.23026

Zaccaroni, M., Passilongo, D., Bucianti, A., Dessi-Fulgheri, F., Facchini, C., Gazzola, A., ... Apollonio, M. (2012). Group specific vocal signature in free-ranging wolf pack. *Ethology Ecology & Evolution*, *24*, 322–331. doi: https://doi.org/10.1080/03949370.2012.664569

Zhang, K., Song, R., Tu, R., Edlund, J., Beskow, J., & Ekström, A. G. (2024, June). Modeling, synthesis and 3D printing of tube vocal tract models with a codeless graphical user interface. In *Proceedings from fonetik 2024* (pp. 155–160). Stockholm, Sweden. doi: https://doi.org/10.5281/zenodo.11396112