

The Fluid Dynamic Origin of Mammalian Ultrasonic Vocalizations

Matthew Dornfeld,* Diego Laplagne, Martin Costa, and Marcelo Magnasco
Rockefeller University
 (Dated: June 18, 2014)

We investigate the physical production mechanism behind mammalian, narrowband, ultrasonic vocalizations- using Long Evans rats as a model organism. We develop a curve fitting algorithm that extracts discontinuous jumps in frequency from many recorded calls from several rats ($n=11$). We also develop a clustering algorithm that fits this data to a fluid dynamic whistle model. By doing so we are able to conclude that the source of the vocalizations is a hole tone type whistle- not vibrating folds, and the frequency jumps are the result of an exchange in stability of flows of the whistle.

Although the purposes of animal vocalizations are numerous, the apparatus by which they are produced seems to be largely conserved across mammalian and bird species. This apparatus consists of membranous vocal fold tissues stretched across the larynx (or syrinx in birds). This tissue attaches to laryngeal muscles, which can alter its tension and degree of adduction. When air is forced past them, by the lungs, the vocal folds oscillate and emit audible sounds, the frequency of which are determined by the vocal fold tension. Across species, the fundamental frequencies of audible mammalian vocalizations, which span several orders of magnitude in mass, approximately obey a power law of the form $f \propto M^{-\alpha}$, where M is the species mass. However, many species of the orders Rodentia and Cetacea are known to produce narrowband ultrasonic whistles and broadband ultrasonic clicks, in addition to their sonic vocalizations. While the sonic vocalizations obey the aforementioned scaling law well, ultrasonic calls from the same species completely diverge from the power law fit by over and order of magnitude (see Fig. 1). There is currently some debate over the exact nature of these ultrasonic calls, and the divergence from the scaling law is indicative that the production mechanism is fundamentally different from that of audible vocalizations [1, 2].

Of the above mentioned ultrasonic vocalizations (USVs), the broadband clicks are believed to serve non-communicative echolocative purposes. In odontocetes, the broadband clicks are known to be produced by a structure, called the phonic lips, which strike the fatty tissue of a structure called the melon, which transmits the vibrations to the surrounding water. However, The production mechanism behind the narrowband calls in cetaceans is currently unknown. Furthermore, they are believed to be directed to conspecifics for the purpose of communication. Thus, the study of these vocalizations has broader implications for the fields of animal communication and ethology as a whole [3]. In rodents, some experimental work has been done, which discounts the possibility the narrowband calls are produced by vibrating vocal folds and indicates the production mechanism is a proper fluid dynamic whistle [4–6]. However, there still does not exist a full quantitative theory of their produc-

tion. It is for these reasons, this paper focuses on the narrowband rodent calls and aims to provide the beginnings of a quantitative, physical, theory of their production by fitting recordings of narrowband vocalizations to a fluid dynamic whistle model. Because of ease of handling long-evans rats are used as a model organism. This analysis is done with the hope that the study of narrowband rodent calls will provide greater insight into those of cetaceans and animal communication in general. Although several

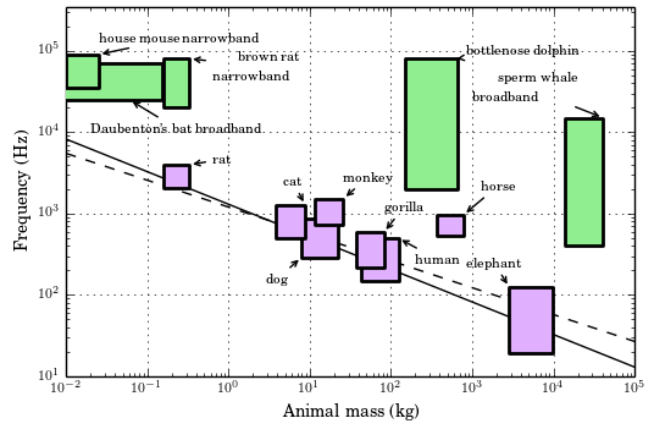


FIG. 1: The purple boxes show the ranges of audible frequencies emitted by several mammals plotted against the ranges of their masses. The solid and dotted lines show plots of a frequency scaling law of the form $f \propto M^{-\alpha}$. The dotted line ($\alpha = \frac{1}{3}$) is a theoretical relationship based on the conjecture that the average frequency emitted in audible vocalizations should be inversely proportional to cube of the linear scale of the animal. The solid line ($\alpha = 0.4$) is a power law fit to the purple boxes [7]. The green boxes show the range of ultrasonic frequencies emitted by several mammalian species plotted against their masses [8–14].

studies indicate rat USVs cannot be caused by vibrating vocal fold tissue and are likely the result of a fluid dynamic whistle, the type of whistle that constitutes this mechanism is still a point of conjecture. In general, fluid dynamic whistles consist of a jet of air, which at higher Reynolds numbers ($Re > 1000$), destabilizes into a vortex street. This vortex street interacts with downstream structures, radiating acoustic energy into the far field,

and inducing a velocity that provides energy to the perturbation frequencies, in the destabilized jet, which are close to the vortex shedding frequency. The perturbation frequencies with more energy become more unstable and are more likely to make contributions to the vortex street. This forms a positive feedback mechanism, which results in a fully formed vortex street with a well defined shedding frequency. The vortex street radiates acoustic energy, at a single frequency, equal to the vortex shedding rate, into the farfield. Analytic results can be obtained by neglecting the complicated destabilization process and assuming the existence of a fully formed vortex street. The equation,

$$\frac{fL}{U_c} = n - \gamma, \quad (1)$$

applies well to all fluid dynamic whistles that do not involve vibration of solid boundaries. The general procedure for it's derivation consists of assuming the form of the vortex street and calculating the velocity induced by the interaction of the vortex street with the downstream structure and matching its phase to that of the perturbations in the jet to ensure the positive feedback mechanism. In equation 1, f is the tonal frequency emitted in the acoustic far field (which is also the vortex shedding frequency), L is a characteristic length scale of the system, U_c is the convection velocity of the vortex street, n is the acoustic mode number, and γ is a constant that is dependent on the type of whistle. The quantity, $\frac{U_c}{f}$ is the spacing between successive vortex rings, thus $n - \gamma$ has the physical interpretation of being the number of disturbance wavelengths that can fit in the length L . The quantity γ is also proportional to the phase difference between the acoustic feedback and the vortex formation rate [15, 16].

It is currently unclear how rat anatomical structures would operate together as a whistle mechanism. Figure 4 shows γ for several different types of whistle mechanisms that can possibly correspond to rat anatomy. Below we will discuss how to fit data extracted from recordings of rat vocalizations to Eq. 1 to get a determination of γ , but first we will discuss how different types of whistles may correspond to rat anatomy. For the rat phonatory system to operate as an edge tone, the vocal folds would focus air coming up from the lungs into a plane jet which impinges on an edge like downstream structure- the most likely candidate for which is the junction between the nasal and buccal cavities. A Krmn vortex street is then shed antisymmetrically into the nasal and buccal cavities [15, 17]. However, if this were the mechanism, it would be expected that sound is emitted equally from both the nose and the mouth, and it has been observed that sound is primarily emitted from the mouth. In addition, the hole formed by the vocal folds have been observed, by camera in anesthetized rats, to be approximately circular in shape throughout vocalization. This precludes the

possibility of them focusing the stream of air into a plane jet [7].

It is also possible air emerging from the vocal folds flows over the junction to the nasal cavity into the buccal cavity, acting as a cavity resonator. Vortices are shed at the leading edge of the opening to the nasal cavity and interact with the cavity and trailing edge to induce a velocity that radiates sound into the far field and provides acoustic feedback to reinforce vortex production at the leading edge. The analysis of cavity resonators, and thus the value of γ , depends on the ratio of the width of the cavity opening to its depth. If this ratio is less than unity, the cavity is considered shallow, and the energy stored in the cavity can be neglected. However, if the ratio is greater than unity, the cavity is considered deep, and this energy must be considered. The length of the nasal cavity is much greater than the width of the entrance to the nasal cavity at the nasal-buccal junction. Thus, the deep cavity model seems to correspond better to rat anatomy. However, for completeness we include both values of γ in 4 [7, 15].

The hole is tone what we believe to be the most biologically realistic model for our system. It consists of a circular jet impinging on a downstream aperture of slightly greater width. In this model, the vocal folds focus the air from lungs into a circular jet. The most likely anatomical structures for the downstream aperture are either the space between the base of the tongue and the soft palate or the one between the epiglottis and the cranial wall of the oropharynx. The circular jet destabilizes into a street of vortex rings axisymmetric about the jet axis. The vortex street causes air in the downstream aperture to oscillate, which radiates sound out of the mouth. The value of γ in Figure 4 was obtained experimentally as well as theoretically under the condition that the length between the apertures is much greater than the radius of the jet- a condition that is not necessarily met in the rat vocal tract as the radius of the opening in the vocal folds during vocalization as well as the distance between the apertures are both on the order of 1 mm [7, 15, 18].

To fit data to this model, we examined the frequency jump points from thousands of recorded calls from eleven different rats (nine males and two females). These recordings were obtained by sampling calls at 250-300 kHz using a condenser ultrasound Avisoft-Bioacoustics CM16/CPMA-5V microphone connected to a National Instruments data acquisition card. The recordings were then divided into call snippets based on long quiescence times between vocalizations. Afterwards, the reassigned spectrogram of each call snippet was computed. Since each vocalizations is largely monotonal, a curve can be fitted to the call in time-frequency space by finding the frequency with the most power and computing the average of the seven frequency points surrounding and including that maximum- weighted by the amount of power in each frequency (see Figure 2). This is done for each time

point. Jump points can then be determined by finding points where the curve rapidly changes value. For each jump point, the after frequency can be plotted against the before. The data naturally groups itself into clusters along straight lines. In doing so, we have extracted features with drastically lower dimensionality than the original data, and we hypothesize that each cluster corresponds to one type of mode transition, with $n \rightarrow n \pm 1$. To fit the extracted jump points to Eq. 1 we developed

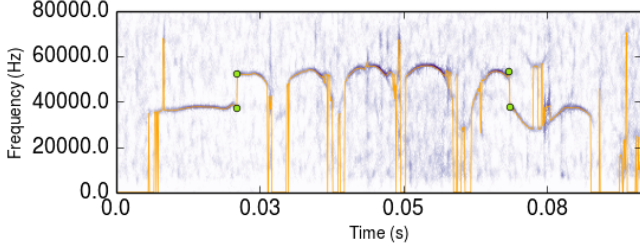


FIG. 2: Reassigned spectrogram of a segment of a rat USV. The orange line is the output of the curve fitting algorithm. The green circles show the output of the jump extraction algorithm.

a clustering algorithm, which as input, accepts a list of slopes of straight lines. The number of clusters in the output is equal to the length of the list. The algorithm then iterates over each jump point in the dataset and computes the perpendicular distance of the point to the line defined by each slope. The point is then assigned to the line for which this value is minimum. If (f_1, f_2) and $(n, n \pm 1)$ are the frequencies and mode numbers before and after a jump respectively, by dividing the after jump frequency mode relation by the before jump relation, we get

$$\frac{f_2}{f_1} = m(c, \gamma) = \frac{n \pm 1 - \gamma}{n - \gamma}, \quad (2)$$

where c is the cluster to which that jump has been assigned, and $(n, n \pm 1)$ is a function of the cluster. Several mode transitions were experimented with to best fit the data. In the extracted jumps, we found transitions in mode numbers ranging from $n = 2$ to $n = 6$. The $n = 1$ mode seems to correspond to 22 kHz alarm calls, which do not exhibit jumps as often and therefore are not represented in our dataset. Not all rats in our study produced enough jumps in all mode regions to fit the data to our algorithm, so underpopulated modes were excluded for some rats. Equation 2 can then be used to generate the slopes needed as input for the algorithm.

We can define a cost function that adds up the squares of the perpendicular distances of each point to their as-

signed line.

$$\Delta x_{i_c} = \left| f_{1,i_c} - \frac{f_{1,i_c} + m(c, \gamma) f_{2,i_c}}{m(c, \gamma)^2 + 1} \right| \quad (3)$$

$$\Delta y_{i_c} = \left| f_{2,i_c} - \frac{m(c, \gamma) (f_{1,i_c} + m(c, \gamma) f_{2,i_c})}{m(c, \gamma)^2 + 1} \right|$$

$$C(\gamma) = \sum_c \frac{1}{N_c} \sum_{i_c=1}^{N_c} (\Delta x_{i_c})^2 + (\Delta y_{i_c})^2, \quad (4)$$

where Δx_{i_c} and Δy_{i_c} are the respective horizontal and vertical distances of the i^{th} point in cluster c to its assigned line, and N_c is the number of points in cluster c . By minimizing Eq. ??eq:cost over γ , we can get a determination for which cluster each point belongs to as well as a numerical calculation for the value γ .

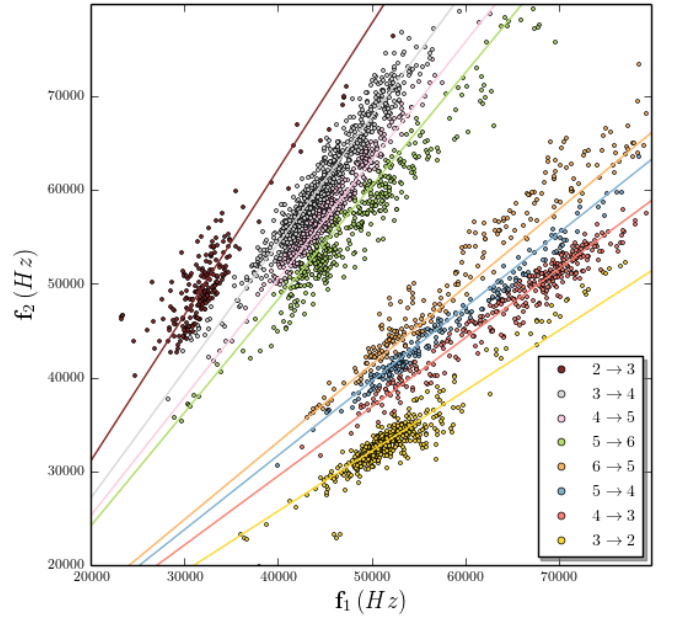


FIG. 3: After jump frequency plotted against before jump frequency for jumps extracted from V4. The points are color coded according to the transition cluster they were assigned by the clustering algorithm. The legend in the bottom right shows which mode transitions ($n \rightarrow n \pm 1$) correspond to the color. The lines are plots of Eq. 2, for $\gamma = \gamma_{min}^{V4}$ and obey the same color coding scheme.

Figure 3 shows the output of the clustering algorithm, for rats V6 (a) and V4 (b), for $\gamma = \gamma_{min}^{rat}$, the values of γ that minimize 4 for a given rat, along with plots, of the lines defined by $m(c, \gamma_{min}^{rat})$, for each cluster. We also experimented with adding a parameter that offset the lines defined by Eq. 2 from the origin. As expected, in all cases, it was found to be effectively zero, so we do not include it in further analysis. For both rats shown in Figure 3, the data is grouped into eight clusters- four

that correspond to up transitions and four that correspond to down transitions. Unlike some other rats used in this study, these ones exhibit significant number of jump points in the $2 \rightarrow 3$ and $3 \rightarrow 2$ transition clusters, which are mostly populated by jumps between constant frequency calls in the 30 kHz range to constant frequency calls in the 50 kHz and vice versa, which were rarely recorded for an unknown reason. These clusters are noticeably distinct from the others and can be separated out by visual inspection. The clusters that represent transitions between higher regions present more of an overlap with each other. The $3 \rightarrow 4$, $5 \rightarrow 6$, $4 \rightarrow 3$, and $6 \rightarrow 5$ clusters show clear drop offs in density at their borders, which seems to be well represented in the output of the algorithm. However, the $4 \rightarrow 5$ and $5 \rightarrow 4$ transition clusters do not seem to be well represented in the data for this rat. Thus, it is hard to identify these clusters based on visual inspection of the density of jump points, although we still believe them to be present in smaller numbers. It is also worth noting that, removing those clusters from the algorithm did not produce a significant change in the calculated value of γ .

The results of Figs. 3 and 4 give clear indications about the nature of the whistle mechanism underlying narrow-band USVs. Upon inspection of Fig 3 evidence of hysteresis can be seen as the densities of opposite transition clusters are not mirror images of each other and seem to be translated and warped along the lines with slope $m(c, \gamma_{min}^{rat})$, which would be expected for a fluid dynamic whistle. Furthermore, the points of each cluster seem to be distributed roughly symmetrically around these lines. This indicates the frequency jumps are the result of a noisy process in the vicinity of a critical point of the system, which would also be expected for a whistle system. Figure 4 shows a comparison of the values of γ_{min} obtained for the rats used in this study with ones known for different fluid dynamic whistles. Based on these results, we can conclusively rule out the edge tone, shallow cavity, and deep cavity mechanisms as being the source of rat USVs. The values of γ obtained from this study correspond best with the one known for the hole tone. Although this correspondence is not exact, we believe it supports our conjecture that the underlying mechanism for USVs is hole tone like. Furthermore, as mentioned above, the value of γ known for the hole tone was obtained under the condition that the length between the up and downstream apertures is much greater than the radius of the jet- a condition that we do not believe is met in the rat vocal tract. In addition, the resonant properties of the upper vocal tract would also affect the value obtained for γ . We believe further theoretical analysis, using these assumption, would produce values γ in agreement our data.

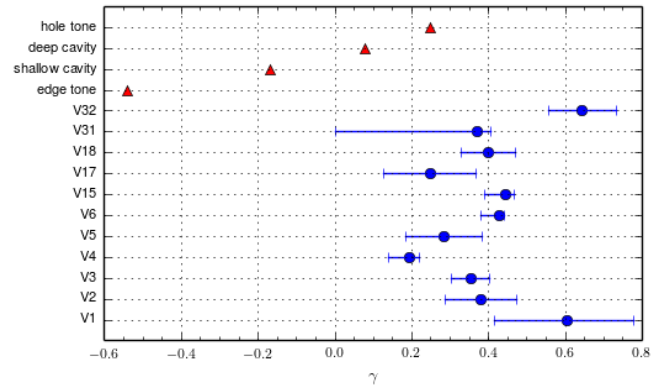


FIG. 4: The blue shows γ_{min}^{rat} along with 95 % confidence interval obtained from bootstrap analysis. The red shows values of γ known for several whistle types [15].

* Electronic address: mdornfel@gmail.com

- [1] N. H. Fletcher, Handbook of Behavioral Neuroscience **19**, 5156 (2010).
- [2] G. S. Berke and J. L. Long, Handbook of Behavioral Neuroscience **19**, 419426 (2010).
- [3] J. S. Reidenberg and J. T. Laitman, Handbook of Behavioral Neuroscience **19**, 451465 (2010).
- [4] T. Riede, Journal of Neurophysiology **106**, 25802592 (2011).
- [5] T. Riede *et al.*, Journal of Biomechanics **44**, 19361940 (2011).
- [6] L. Roberts, Ultrasonics **13**, 8388 (1975).
- [7] S. M. Brudzynski and N. H. Fletcher, Handbook of Behavioral Neuroscience **19**, 6976 (2010).
- [8] N. R. White, M. Prasad, R. J. Barfield, and J. G. Nyby, Physiology & Behavior **63**, 467 (1998).
- [9] R. Berry, Field Studies **3**, 219 (1970).
- [10] M. B. Fenton, C. V. Portfors, I. L. Rautenbach, and J. M. Waterman, Canadian Journal of Zoology **76**, 11741182 (1998).
- [11] G. Jones and E. Teeling, Trends in Ecology & Evolution **21**, 149156 (2006).
- [12] W. Bogdanowicz, Mammalian species 1 (1994).
- [13] A. S. Frankel, Encyclopedia of Marine Mammals 10561071 (2009).
- [14] H. Whitehead, Encyclopedia of Marine Mammals 10911097 (2009).
- [15] M. S. Howe, *Acoustics of Fluid-Structure Interactions (Cambridge Monographs on Mechanics)* (Cambridge University Press, ADDRESS, 2008).
- [16] W. K. Blake and A. Powell, Recent Advances in Aeroacoustics 247325 (1986).
- [17] D. K. Holger, T. A. Wilson, and G. S. Beavers, The Journal of the Acoustical Society of America **62**, 1116 (1977).
- [18] R. C. Chanaud and A. Powell, The Journal of the Acoustical Society of America **37**, 902 (1965).