High-fidelity continuum modeling predicts voiced sound production

Authors: W. Jiang[1,§], J.H. Rasmussen[2,§,¶], Q. Xue[1], M. Ding[3], X. Zheng[1*], C.P.H. Elemans[2*]

Affiliations:

[1] Department of Mechanical Engineering, University of Maine, Orono, ME, USA.

[2] Department of Biology, University of Southern Denmark, Odense M, Denmark.

[3] Department of Health, University of Southern Denmark, Odense M, Denmark.

§ Equal contribution

[¶] Current address: Department of Marine Biology, Texas A&M University at Galveston, Galveston, TX, USA.

*Correspondence to: xudong.zheng@maine.edu and coen@biology.sdu.dk

Classification: Biological sciences - Biophysics and Computational Biology, Neuroscience, Physiology

Keywords: Birdsong, human voice, computational fluid dynamics

Abstract

Voiced sound production is the primary form of acoustic communication in terrestrial vertebrates, particularly birds and mammals, and including humans. Developing a causal physics-based model that links descending vocal motor control to tissue vibration and sound, requires embodied approaches that include realistic representations of voice physiology. Here we first implement and then experimentally test a high-fidelity three-dimensional continuum model for voiced sound production in birds. Driven by individual-based physiologically quantifiable inputs, combined with non-invasive inverse methods for tissue material parameterization, our model accurately predicts observed key vibratory and acoustic performance traits. These results demonstrate that realistic models lead to accurate predictions supporting the continuum model approach as a critical tool towards a causal model of motor control of voiced sound production.

Introduction

During voiced sound production laryngeal vocal folds (VFs) or analogous syringeal structures exhibit self-sustained vibrations that through fluid-tissue interactions and elastic restoring forces mechanically convert expiratory airflow into pulsatile airflow, which causes air pressure oscillations constituting the acoustic excitation of the system (*1–4*). This physical framework is known as the myoelastic-aerodynamic theory (MEAD) (*2*, *3*) and shared between mammals and birds (*5*). In the human larynx, which has been most intensively studied, VF tissue properties, aerodynamic forces and neuromuscular control of laryngeal and VF geometry all interact to ultimately control VF kinematics and glottal flow that set crucial acoustic source parameters such as fundamental frequency, registers, and spectral slope (*1*, *4*, *6*). In contrast to humans, birds are not as well-studied but have the significant advantage that the entire neuromechanical motor control loop of their voice production is experimentally accessible (*7*, *8*). Furthermore the process of song motor sequence learning in songbirds can be monitored individually and shares many parallels with human speech acquisition (*9–12*). Thus birds present a tractable model system for studying the physics and learned motor control of voiced sound production and to inform on basic principles of MEAD systems control.

Causally linking descending motor control to voiced sound production requires computational biophysical models to explore the multidimensional control space (*13*), which the brain also learns to navigate. Particularly high-fidelity continuum models that include the full fluid-structure-acoustics interaction (FSAI) complexity of voiced sound production in anatomically realistic geometries of vocal fold and tract (*4*, *14–19*) are essential to develop when realistic representations of voice physiology and biomechanics are essential, such as in the clinical management of voice disorders (*1*), or understanding motor control of voice (*20*, *21*) or

(bird)song (*22*). However, the continuum model approach has not yet been applied to birds during voiced sound production and in general currently faces six key challenges (*4*, *20*): On the computational side, (*i*) the implementation of complex FSAI models itself capable of resolving the large range of spatial and time scales remains challenging (*14–18*, *20*). On the experimental side, although partial physiological datasets have been collected in mammals (*1*, *23*, *24*), we currently lack quantification of all critical physiological data in single individuals, including (*ii*) anatomical and geometrical boundary conditions, (*iii*) parameterization of VF tissue properties, and (iv) flow, pressure and VF posturing changes due to motor control, while (v) quantifying the time-resolved VF shape within single oscillations. Obtaining these complete individual-based datasets would lift the last critical roadblock and (vi) allow thorough experimental testing of FSAI model predictions accuracy to strengthen confidence in the continuum model approach.

Results

Here we meet the above challenges in an avian model system and implemented and experimentally test predictions of a physics-based, high-fidelity three-dimensional (3D) continuum model of voiced sound production. First, we implemented a 3D immersed-boundary (IBM), finite-element method based fluid-structure interaction solver (*14*, *15*, *17*, *18*) (See methods), where airflow was governed by 3D, unsteady, viscous, incompressible Navier–Stokes equations and dynamics of the avian VF analog were governed by the Navier equation. The fluid and structure solvers were explicitly coupled through a Lagrangian interface where airway and vibrating structures contacted. To allow rapid simulations when resolving the large range of spatial and time scales in simulating voice production, we used the Message-Passing-Interface parallelization method (*25*), combined with IBM as an advanced numerical method designed

80 especially for simulating moving boundaries. By carrying out simulations on Cartesian grids, IBM eliminates the need for complicated re-meshing algorithms in conventional body-fitted mesh methods, and can deal with complex moving boundaries.

To address the second challenge, i.e., anatomical and geometrical boundary conditions, we focused on rock pigeons, where we previously (*5*, *26*) achieved high-quality kinematic data
85 of the avian VF analog, the lateral vibratory masses (LVMs, Fig. 1A). We quantified the anatomical and geometrical boundary conditions for five individuals using DiceCT scans and implemented high-resolution finite-element meshes of the syrinx (Fig. 1A). To mimic the anatomical constraint of the trachea rings on the motion of the LVMs, zero displacement boundary conditions were applied at mesh nodes where tracheal rings were located. Unilateral
90 LVMs consisted of 1,988-4,620 and 2,159-4,693 nodes and 8,101-18,657 and 9,838-19,131 elements in the solid domain for left and right LVM respectively (Table S1).

To address the third challenge, i.e., parameterization of VF tissue properties, we developed a novel non-invasive, combined experimental/simulation approach to parameterize tissue elasticity for each individual. We experimentally induced static LVM displacement by
95 stepwise increase of the transmural pressure (Fig. 1B). We subsequently used the LVM finite-element model combined with genetic algorithm based optimization (*27*) to simulate LVM displacement as a function of transmural pressure (Fig 1C) and find the elastic modulus (EM) value with the smallest difference between experiment and simulation (Fig 1C). These values were used for further dynamic simulation and ranged from 1.8 to 4.0 kPa for the five individuals
100 (Table S1).

For dynamic performance, we used a blinded procedure where the modeling team (WJ, QX, XZ) had access only to geometry, static loading test and pressure boundary conditions, but

was blind to all other data of the experimental procedures (performed by JHR, CPHE). To address the fourth challenge, i.e., quantifying flow, pressure and VF posturing changes due to motor control, we applied boundary conditions where LVM vibration was obtained reliably previously (*5*, *26*) to both the experimental preparation and FSAI model (p_b = 1.0 kPa, p_{icas} = 0.5 kPa). Both the *in vitro* syrinx and corresponding individual FSAI models demonstrated self-sustained stable oscillations in all five cases (Fig 2). To address the fifth challenge, quantifying the time-resolved VF shape within oscillations, we took advantage of both the unique coronal view offered by the pigeon syrinx (*5*) and the lack of a dorsoventral vibrational component (*26*), to quantify the time-resolved syringeal or glottal opening as a function of caudo-cranial position, i.e. a coronal glottovibrogram (cGVG, Fig 2C).

To rigorously test the predictions of the FSAI model (thus meeting the sixth and final challenge), we compared key vibratory and acoustic predictions by the simulations (SIM) to the behavior in the experiment (EXP). The cGVG allowed for comparing four key parameters describing vibratory kinematics: fundamental frequency of the vibration (f_o), the speed of the caudocranial (CC) wave, peak of the minimal glottal opening, and the open-closed quotient (OCQ). In EXP, f_o of the LVM vibration measured 123±25 Hz (median: 114 Hz; range: 95-154 Hz, N=5), CC speed was 0.82±0.34 m/s (median: 0.88 m/s; range: 0.48-1.27 m/s, N=5), peak minimal opening was 1.80±0.24 mm (median: 1.80 mm; range: 1.49-2.11 mm; N=5) and OCQ was 1.3±.0.7 (median: 1.1, range: 0.7-2.5; N=5). In SIM, f_o of the LVM vibration measured 143±18 Hz (median: 152 Hz; range: 115-158 Hz, N=5), CC speed was 1.02±0.26 m/s (median: 1.03 m/s; range: 0.68-1.38 m/s, N=5), peak minimal opening was 1.06±0.43 mm (median: 0.90 mm; range: 0.62-1.64 mm; N=5) and OCQ was 1.6±0.3 (median: 1.6, range: 1.1-2.1; N=5). At group level, all predicted values were not significantly different from the experimental values

with p=0.18, 0.50, 0.06 and 0.11 respectively (Heteroscedastic two-tailed paired t-test, N=5) (Fig 2D). We furthermore compared the time-resolved LVM shape between EXP and SIM at fixed phases within an oscillatory cycle (Fig 3). During the closed phase (0-120 deg) the simulated LVM shape matched EXP very well and was never significantly different from the bootstrapped shape ($p<0.01$, see methods). The only observed significant ($p>0.05$, 2-sample Kolmogorov-Smirnov test) discrepancy occurred during late closed/early opening in three animals on one side, where the LVM mass tended to move ~0.5 mm more cranial (superior in human anatomical terminology) in EXP compared to SIM (Fig 3BC). Taken together, our FSAI model accurately predicted key kinematic parameters of LVM motion.

Lastly, we compared SIM prediction and EXP observation of two additional key acoustic parameters specifying a sound source – in addition to f_o (Fig 2D) -, namely source level and spectral slope and these did not differ significantly (Fig 4A). Because the simulated LVM vibratory kinematics matched our observations, we exploited our FSAI models to calculate parameters that could not be quantified in the current experiments, such as spatiotemporal pressure and flow velocity distributions over the vibratory cycle (Fig 4B-E). The convergent LVM shape during opening causes high glottal pressures (~0.9 kPa) that transfer 23.4±15.7 μJ of positive energy (N=5) from flow into LVM (Fig 4F), facilitating opening. When maximum opening is reached at 300° phase, the LVMs are straight (0° angle in Fig 4E) reducing glottal pressure. During early closing the inferior LVM edge is moving inwards causing an energy transfer of -7.2±2.2 μJ back into the flow. Consecutively, the LVMs close the glottis by moving together in a divergent shape causing rapid pressure reduction driven by elastic forces (Arrow in Fig 4E). Furthermore, flow inertia in the trachea causes negative pressures (-0.52±0.23 kPa) near the glottis exit prior to full closure (blue region in Fig 4D), facilitating closing. Interestingly,

another positive energy transfer to LVMs is observed nearly the end of the cycle. Our avian data thus show that two primary factors contribute to the pressure asymmetry during VF opening and closing that drive self-sustained oscillation: 1) an alternating convergent/divergent medial surface profile and 2) airflow inertia, corroborating earlier model predictions (*28*) and measurements on (hemi)larynges (*29*) in mammalian voice production.

Discussion

We faced the presently outstanding challenges for the continuum model approach in a comparative context and implemented a physics-based, 3D, high-fidelity continuum model of voiced sound production in birds. Driven by physiological parameters (static geometry, tissue elasticity and boundary conditions) and without optimization of either geometry or material properties parameterization on dynamic performance (*30*) our FSAI model accurately predicted key performance traits of tissue motion and acoustics. While the FSAI model implementation itself is complex (*14–18*, *20*), the inputs are simple and have, most importantly, directly measurable physiological, material and geometrical properties. Measurements of VF material properties (*31*, *32*), initial configuration, initial stress and detailed flow-induced 3D VF motion (*33*) have been achieved separately in human and mammalian model systems, but complete physiological data sets have not been obtained in these clades, nor in any birds, in single individuals nor consecutively used to thoroughly test FASI model predictions in a blinded approach. Recent studies encouragingly suggested that realistic continuum 3D models lead to more robust VF dynamics compared to 1D and 2D models (*34*), and our data shows that they also lead to accurate predictions performance. Our results strongly support the continuum model approach as a critical step in integrating experimental *in vitro, ex vivo and in vivo* data with brute

force computational approaches towards a causal model of motor control of voiced sound production applicable to birds and mammals (*4*, *20*).

Continuum models currently allow only limited systematic parameters exploration because they are computationally expensive (*20*). In contrast, phenomenological reduced order (RO) models - that simplify VFs to one or two masses (*35*, *36*)- allow broad exploration of a simplified control space with less computational power and as such have been used widely to capture the dynamics of mammalian (*37*) and avian voiced sound production (*13*, *38–40*). However, because RO models are not first-principle physics models they essentially lack physiological representation of tissue properties and geometry (*1*, *4*, *37*) and care has to be taken to applying them if realistic representations of voice physiology are essential (*1*). Reducing computational costs and time is thus crucial for making continuum models useful in experimental or clinical settings (*1*). Excitingly, rapid developments in machine learning tools combined with supercomputing provide promising avenues to predict complex flow solvers (*41*, *42*) that could significantly reduce computational power and allow elaborate parameter exploration.

Because animal behaviors result from complex system-wide interactions between nervous system, body, and surrounding environment, the activity of neural circuits controlling sound can only be understood by considering the biomechanics of sound generation, muscles, bodies, and the exterior world (*43–46*). In vocal behaviors, indeed small tissue changes can lead to state changes in vocal dynamics (*47*) and changes in muscle performance interact with neural coding in songbirds (*48*). An embodied approach to voice production as presented here that captures accurate parameterization of vocal organs is essential to further our understanding how underlying neural mechanisms and biomechanics interact to drive vocal behavior, from direct

195 motor control to postnatal vocal developmental, and changes on comparative and evolutionary scale. Additionally, because most laryngeal human voice disorders, such as VF lesions or aberrant motor control of VF tensioning are caused by changes in VF geometry, structural integrity, or kinematics (*4*), embodied models have direct clinical relevance to relate pathological physical behaviors with abnormal vocal output (*1*) and are predicted to bring patient specific
200 model-assisted phonosurgery realistically close within a decade (*49*, *50*).

References and notes:

1. Z. Zhang, Mechanics of human voice production and control. *J. Acoust. Soc. Am.* **140**, 2614–2635 (2016).

2. I. R. Titze, *Principles of Voice Production* (National Center for Voice and Speech, Denver, 2000).

3. J. van den Berg, Myoelastic-aerodynamic theory of voice production. *J. Speech Hear. Res.* **1**, 227–244 (1958).

4. R. Mittal, B. D. Erath, M. W. Plesniak, Fluid Dynamics of Human Phonation and Speech. *Annu. Rev. Fluid Mech.* **45**, 437–467 (2013).

5. C. P. H. Elemans, J. H. Rasmussen, C. T. Herbst, D. N. Düring, S. A. Zollinger, H. Brumm, K. Srivastava, N. Svane, M. Ding, O. N. Larsen, S. J. Sober, J. G. Švec, Universal mechanisms of sound production and control in birds and mammals. *Nat. Commun.* **6** (2015), doi:10.1038/ncomms9978.

6. E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. C. Guiod, S. L. Goldman, Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *J. Speech Hear. Res.* **38**, 1212–1223 (1995).

7. T. Riede, F. Goller, Peripheral mechanisms for vocal production in birds - differences and similarities to human speech and singing. *Brain Lang.* **115**, 69–80 (2010).

8. C. P. H. Elemans, The singer and the song: The neuromechanics of avian sound production. *Curr. Opin. Neurobiol.* **28**, 172–178 (2014).

9. M. S. Brainard, A. J. Doupe, Translating Birdsong: Songbirds as a Model for Basic and Applied Medical Research. *Annu. Rev. Neurosci.* **36**, 489–517 (2013).

10. O. Tchernichovski, P. P. Mitra, T. Lints, F. Nottebohm, Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science.* **291**, 2564–2569 (2001).

11. M. A. Long, M. S. Fee, Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*. **456**, 189–194 (2008).

12. T. S. Okubo, E. L. Mackevicius, H. L. Payne, G. F. Lynch, M. S. Fee, Growth and splitting of neural sequences in songbird vocal development. *Nature*. **528**, 352–357 (2015).

13. A. Amador, Y. S. Perl, G. B. Mindlin, D. Margoliash, Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature*. **495**, 59–64 (2013).

14. Q. Xue, R. Mittal, X. Zheng, S. Bielamowicz, Computational modeling of phonatory dynamics in a tubular three-dimensional model of the human larynx. *J. Acoust. Soc. Am.* **132**, 1602–1613 (2012).

15. Q. Xue, X. Zheng, R. Mittal, S. Bielamowicz, Subject-specific computational modeling of human phonation. *J. Acoust. Soc. Am.* **135**, 1445–1456 (2014).

16. W. Jiang, X. Zheng, Q. Xue, Computational modeling of fluid-structure-acoustics interaction during voice production. *Front. Bioeng. Biotechnol.* **5** (2017), doi:10.3389/fbioe.2017.00007.

17. X. Zheng, Q. Xue, R. Mittal, S. Beilamowicz, A coupled sharp-interface immersed boundary-finite-element method for flow-structure interaction with application to human phonation. *J. Biomech. Eng.* **132** (2010), doi:10.1115/1.4002587.

18. R. Mittal, H. Dong, M. Bozkurttas, F. M. Najjar, A. Vargas, A. von Loebbecke, A versatile sharp interface immersed boundary method for incompressible flows with complex boundaries. *J. Comput. Phys.* **227**, 4825–4852 (2008).

19. F. Alipour, D. A. Berry, I. R. Titze, A finite-element model of vocal-fold vibration. *J. Acoust. Soc. Am.* **108**, 3003–3012 (2000).

20. Z. Zhang, Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *J. Acoust. Soc. Am.* **139**, 1493–1507 (2016).

21. E. J. Hunter, I. R. Titze, F. Alipour, A three-dimensional model of vocal fold abduction/adduction. *J. Acoust. Soc. Am.* **115**, 1747–1759 (2004).

22. E. P. Kingsley, C. M. Eliason, T. Riede, Z. Li, T. W. Hiscock, M. Farnsworth, S. L. Thomson, F. Goller, C. J. Tabin, J. A. Clarke, Identity and novelty in the avian syrinx. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 10209–10217 (2018).

23. D. K. Chhetri, J. Neubauer, J. L. Bergeron, E. Sofer, K. A. Peng, N. Jamal, Effects of asymmetric superior laryngeal nerve stimulation on glottic posture, acoustics, vibration. *Laryngoscope*. **123**, 3110–3116 (2013).

24. D. K. Chhetri, J. Neubauer, D. A. Berry, Neuromuscular control of fundamental frequency and glottal posture at phonation onset. *J. Acoust. Soc. Am.* **131**, 1401–1412 (2012).

25. W. Gropp, E. Lusk, N. Doss, A. Skjellum, A high-performance, portable implementation of the MPI message passing interface standard. *Parallel Comput.* **22**, 789–828 (1996).

26. J. H. Rasmussen, C. T. Herbst, C. P. H. Elemans, Quantifying syringeal dynamics in vitro using electroglottography. *J. Exp. Biol.* **221** (2018), doi:10.1242/jeb.172247.

27. G. Liu, B. Geng, X. Zheng, Q. Xue, H. Dong, G. V. Lauder, An image-guided computational approach to inversely determine in vivo material properties and model flow-structure interactions of fish fins. *J. Comput. Phys.* **392**, 578–593 (2019).

28. I. R. Titze, The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.* **83**, 1536–1552 (1988).

29. F. Alipour, S. Jaiswal, Phonatory characteristics of excised pig, sheep, and cow larynges. *J. Acoust. Soc. Am.* **123**, 4572–4581 (2008).

30. S. Chang, C. K. Novaleski, T. Kojima, M. Mizuta, H. Luo, B. Rousseau, Subject-Specific Computational Modeling of Evoked Rabbit Phonation. *J. Biomech. Eng.* **138** (2016), doi:10.1115/1.4032057.

31. G. R. Dion, S. Jeswani, S. Roof, M. Fritz, P. G. Coelho, M. Sobieraj, M. R. Amin, R. C. Branski, Functional assessment of the ex vivo vocal folds through biomechanical testing: A review. *Mater. Sci. Eng. C*. **64**, 444–453 (2016).

32. A. K. Miri, Mechanical characterization of vocal fold tissue: A review study. *J. Voice*. **28**, 657–667 (2014).

33. A. M. Vahabzadeh-Hagh, Z. Zhang, D. K. Chhetri, Three-dimensional posture changes of the vocal fold from paired intrinsic laryngeal muscles. *Laryngoscope*. **127**, 656–664 (2017).

34. X. Zheng, R. Mittal, Q. Xue, S. Bielamowicz, Direct-numerical simulation of the glottal jet and vocal-fold dynamics in a three-dimensional laryngeal model. *J. Acoust. Soc. Am.* **130**, 404–415 (2011).

35. K. Ishizaka, J. L. Flanagan, Synthesis of Voiced Sounds From a Two□Mass Model of the Vocal Cords. *Bell*

Syst. Tech. J. **51**, 1233–1268 (1972).

36. B. H. Story, I. R. Titze, Voice simulation with a body-cover model of the vocal folds. *J. Acoust. Soc. Am.* **97**, 1249–1260 (1995).

37. B. D. Erath, M. Zañartu, K. C. Stewart, M. W. Plesniak, D. E. Sommer, S. D. Peterson, A review of lumped-element models of voiced speech. *Speech Commun.* **55**, 667–690 (2013).

38. G. B. Mindlin, The physics of birdsong production. *Contemp. Phys.* **54**, 91–96 (2013).

39. R. Zaccarelli, C. P. H. Elemans, W. T. Fitch, H. Herzel, Modelling bird songs: Voice onset, overtones and registers. *Acta Acust. united with Acust.* **92**, 741–748 (2006).

40. C. P. H. Elemans, R. Zaccarelli, H. Herzel, Biomechanics and control of vocalization in a non-songbird. *J. R. Soc. Interface*. **5**, 691–703 (2008).

41. B. Kim, V. C. Azevedo, N. Thuerey, T. Kim, M. Gross, B. Solenthaler, Deep Fluids: A Generative Network for Parameterized Fluid Simulations. *Comput. Graph. Forum*. **38**, 59–70 (2019).

42. B. N. Hanna, N. T. Dinh, R. W. Youngblood, I. A. Bolotnov, Machine-learning based error prediction approach for coarse-grid Computational Fluid Dynamics (CG-CFD). *Prog. Nucl. Energy*. **118**, 103140 (2020).

43. E. D. Tytell, P. Holmes, A. H. Cohen, Spikes alone do not behavior make: Why neuroscience needs biomechanics. *Curr. Opin. Neurobiol.* **21**, 816–822 (2011).

44. D. N. Düring, C. P. H. Elemans, in *Vertebrate Sound Production and Acoustic Communication* (2016), pp. 119–157.

45. Y. S. Zhang, A. A. Ghazanfar, Vocal development through morphological computation. *PLoS Biol.* **16** (2018), doi:10.1371/journal.pbio.2003933.

46. K. Nishikawa, A. A. Biewener, P. Aerts, A. N. Ahn, H. J. Chiel, M. A. Daley, T. L. Daniel, R. J. Full, M. E. Hale, T. L. Hedrick, A. K. Lappin, T. R. Nichols, R. D. Quinn, R. A. Satterlie, B. Szymik, Neuromechanics: An integrative approach for understanding motor control. *Integr. Comp. Biol.* **47**, 16–54 (2007).

47. Asif C. P. H. Elemans, Vocal Motor Performance in Birdsong Requires Brain-Body Interaction. *eNeuro*. **6** (2019), doi:10.1523/ENEURO.0053-19.2019.

48. I. Adam, C. P. H. Elemans, Vocal Motor Performance in Birdsong Requires Brain-Body Interaction. *eNeuro*. **6** (2019), doi:10.1523/ENEURO.0053-19.2019.

49. R. Mittal, X. Zheng, R. Bhardwaj, J. H. Seo, Q. Xue, S. Bielamowicz, Toward a simulation-based tool for the treatment of vocal fold paralysis. *Front. Physiol.* **MAY** (2011), doi:10.3389/fphys.2011.00019.

50. H. Sadeghi, S. Kniesburges, S. Falk, M. Kaltenbacher, A. Schützenberger, M. Döllinger, Towards a clinically applicable computational larynx model. *Appl. Sci.* **9** (2019), doi:10.3390/app9112288.

51. B. D. Metscher, Micro CT for comparative morphology: Simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. *BMC Physiol.* **9** (2009), doi:10.1186/1472-6793-9-11.

52. D. N. Düring, A. Ziegler, C. K. Thompson, A. Ziegler, C. Faber, J. Müller, C. Scharff, C. P. H. Elemans, The songbird syrinx morphome: A three-dimensional, high-resolution, interactive morphological map of the zebra finch vocal organ. *BMC Biol.* **11** (2013), doi:10.1186/1741-7007-11-1.

53. J. Seo, R. Mittal, A high-order immersed boundary method for acoustic wave scattering and low-Mach number flow-induced sound in complex geometries. *J. Comput. Phys.* **230**, 1000–1019 (2011).

54. J. H. Seo, Y. J. Moon, Linearized perturbed compressible equations for low Mach number aeroacoustics. *J. Comput. Phys.* **218**, 702–719 (2006).

55. C. P. H. Elemans, *How do birds sing? Sound analysis-mechanical modelling-muscular control* (Wageningen University, 2004).

Acknowledgements: The authors with to thank T Christensen, S Jakobsen, P Martensen and F Mortensen for technical support, I Adam for discussion, and C Herbst and J Rattcliffe for comments on the manuscript. **Funding**: This research was supported by the Danish Research Council and NovoNordisk Foundation (CPHE). **Author contributions**: WJ, JHR, QX, XZ, CPHE designed research; WJ, JHR, MD, QX, XZ, CPHE performed research, with the modeling team consisting of WJ, QX and XZ, and experimental team of JHR and CPHE; CT scans were performed by JHR and MD; QX, XZ, CE contributed new reagents or analytic tools. WJ, JHR, QX, XZ, CPHE analyzed data; CPHE wrote the first draft; All authors read and approved the final manuscript. **Competing interests:** The authors declare no competing financial interests. **Competing interests:** None declared. **Data and materials availability:** All data and code used for analysis are available upon request.

Supplementary Materials

Materials and Methods

Table S1

Fig S1

MATERIALS AND METHODS

Subjects

345 We used 6 adult domestic pigeons (*Columba livia*) obtained from local breeders. The subjects were kept in an outdoor aviary (3x6x2 meters) at the University of Southern Denmark with access to food and water *ad libitum*. All procedures were carried out in accordance with the Danish Animal Experiments Inspectorate (Copenhagen, Denmark).

350 Experimental Setup

Subjects were euthanized by overdosing with Isoflurane (Baxter Medicals, Il, USA) and the syrinx was extracted from the primary bronchi up to including ~10 cm of trachea. The syrinx was mounted to tracheal and bronchial connectors in the experimental setup described in detail in (*5*, *26*). In brief, this setup allows fine-control of bronchial (p_b) and air sac (p_{as}) pressure, flow, 355 temperature, and simultaneous high-speed imaging from sagittal and endoscopic tracheal views. Both pressures were controlled with a dual valve differential pressure PID controller (0–10 kPa, model PCD, Alicat Scientific, AZ, USA), referenced to atmospheric pressure. Bronchial mass flow was measured with flow sensors (0-3 liters/min, PMF 2104, Posifa Microsystems, San Jose, USA). Sound was recorded with a ½-inch pressure microphone (model 46AD with preamplifier 360 type 26AH, G.R.A.S., Denmark), amplified and high-pass filtered (10 Hz, 3-pole Butterworth filter, model 12AQ, G.R.A.S.). The microphone sensitivity was measured before each experiment (sound calibrator model 42AB, G.R.A.S.). The microphone was placed 5-10 cm from the tracheal connector outlet on a 45° angle to avoid air jets from the tracheal outlet. Microphone, pressure and flow signals were low-pass filtered (10 kHz, EF120, Thor labs, NJ, 365 USA) and digitized at 48 kHz (USB 6259, 16 bit, National Instruments, Austin, TX, USA)

together with TTL trigger to synchronization for the camera systems. All control and analysis software was written in Matlab (Mathworks).

Experimental Procedures

370 *1. LVM tissue properties*: After mounting we first measured the stress-stain relationship of the LVM tissue using a static loading test. We imaged the static deformation of the LVMs using a 1.2 mm flexible endoscope (Scholly, Germany) attached to a videokymographic (VKG) system (Videokymographic camera 2156, Cymo B.V., The Netherlands), which combines a high-speed linescan camera (7,200 lines/sec) with a full frame CMOS camera (25 frames/sec), while 375 stepwise increasing transmural pressure every 1.5 s from 0-2 kPa (in increments of 0.1 kPa) corresponding to the *in vivo* pressure range (Beckers, 2003; Gaunt, Gaunt, & Casey, 1982; Elemans 2008). To prevent airflow through the syrinx, the trachea was closed. The analog video output from the VKG was digitized together with the microphone and a synchronization signal using a video capturing device (Intensity Extreme, Black Magic Design, Australia). We 380 constructed digital kymograms perpendicular across the LVM. The edge position was quantified using outwards-in threshold detection of the edge. The length of dorsoventral glottis at mid LVM position was determined from each CT scan and allowed us to calibrate LVM displacement from the DKG in mm. For statistical purpose, five 5 technical replicates were made for each transmural pressure setting.

385

2. LVM kinematics during voiced sound production. Second, we quantified the time-resolved motion of the LVMs, flow, pressure and sound during phonation. We applied the boundary conditions of p_b = 1.0 kPa, p_{as} = 0.5 kPa, where LVM oscillations reliably occurred (*5*, *26*). The

resulting positive transmural pressure ($p_t = p_b - p_{as}$) mimics activation of the TL muscle (). A
390 powerful, stable light source (1700 Lumen white LED powered by PS23023, HQ Power amplifier, Belgium) trans-illuminated the syrinx from dorsal allowed us to capture LVM motion with a high-speed camera (MotionPro-X4, 12 bit CMOS sensor, Integrated Design Tools, Inc.; 4,000 frames/sec) mounted on a stereomicroscope (M165-FC, Leica Microsystems). Before and after experiments we made a photo of the mounted preparation with a Leica DFC400 digital
395 camera mounted on the M165-FC stereoscope. LVM shape was hand-traced in all specimens on 150-200 selected consecutive high-speed images (covering at least 5 full vibratory cycles) using a custom-made Matlab graphical user interface.

3. *Syrinx geometry based on DiceCT scan*. Lastly, after experiments were completed we
400 mounted the syrinx with Austerlitz insect pins (Fine Science Tools, Germany) onto a 5 mm thick sheet of stiff silicon, taking care to mimic the exact geometry of the syrinx during the experiments based on earlier photos. The preparation was submerged into phosphate buffered saline solution (PBS) with 4% paraformaldehyde (PFA) for at least 24 hours to fix the geometry of each preparation. After fixation, the preparations were transferred to PBS and kept at 5C.
405 After mounting, after fixation and prior to scanning we made photos with a Leica DFC400 digital camera mounted on the M165-FC stereoscope to quantify shrinkage and shape changes. One preparation was omitted for further analysis as it twisted after fixation compared to the experimental situation. Preparations were removed from the silicon base and stained in 10% IKI in demineralized water and rotated at 5 RPM for 24 hours twice (*51, 52*). Preparations were
410 washed in demineralized water 8 times just prior to being transferred into a microCT scanner (vivaCT 40, Scanco Medical A., Brüttisellen, Switzerland). Scanning settings were 70 kV and 57

μA. After scanning, the 3D reconstruction cubic voxel size measured 10.5*10.5*10.5 μm3 (2048*2048*2048 pixels) with 32-bit gray-levels.

Blinding Procedure and Data Disclosure

To eliminate the chance of fitting model predictions on experimental data, we implemented the following blinding procedure. The experimental team (JHR, CPHE) performed the experimental procedures listed above. Afterwards, the modeling team (WJ, QX, XZ) was given access only to the DiceCT scan, static loading test (LVM displacement and transmural pressure) and asked to predict key performance traits of LVM kinematics (glottovibrogram) and acoustic waveforms during voiced sound production under listed posturing and pressure boundary conditions. After the simulations of all preparations were completed all experimental data was disclosed and analyzed.

Fluid-Structure Acoustic Interaction (FSAI) Model

The computational solver was built upon an immersed-boundary-finite-element method based fluid-structure interaction solver (*14*, *15*, *17*, *18*). The airflow was governed by the three-dimensional, unsteady, viscous, incompressible Navier–Stokes equations:

$$\nabla \cdot \vec{U} = 0$$

$$\frac{\partial \vec{U}}{\partial t} + \left(\vec{U} \cdot \nabla\right)\vec{U} = -\frac{1}{\rho_0}\nabla \mathrm{P} + \upsilon_0 \nabla^2 \vec{U} \tag{1}$$

where $\vec{U}$, ρ_0, P, υ_0 are the incompressible flow velocity, density, pressure and kinematic viscosity, respectively. The LVM dynamics was governed by the Navier equation with a linear stress-strain relationship:

$$\rho_{tiss}\frac{\partial^2\vec{d}}{\partial t^2} = \bar{\bar{\sigma}}\cdot\overleftarrow{\nabla} + \rho_{tiss}\vec{f} \tag{2}$$

where ρ_{tiss} is the tissue density, $\vec{d}$ is the displacement, $\bar{\bar{\sigma}}$ is the stress tensor, and $\vec{f}$ is the body force. Details regarding the numerical algorithm of the flow and solid solvers can be found in (*17*).

The fluid and structure solvers were explicitly coupled through a Lagrangian interface where airway and LVMs contact. In each iteration, the fluid solver was first marched by one step with the existing deformed shape and velocities of the LVM surface as the boundary conditions. The forces at the LVM surface were then calculated with the new fluid pressure. At the second step, the solid solver was marched by one step with the updated surface traction. The deformation and velocities of the LVM surface formed the new fluid boundary condition, so that the fluid solver and solid solver were coupled. The acoustic pressure was first calculated by the linearized perturbed compressible equation (LPCE) based on the hydrodynamic/acoustic splitting method (*53, 54*):

$$\begin{gathered}\frac{\partial\rho'}{\partial t} + \left(\vec{U}\cdot\nabla\right)\rho' + \rho_0(\nabla\cdot\vec{u}') = 0\\ \frac{\partial\vec{u}'}{\partial t} + \nabla\left(\vec{u}'\cdot\vec{U}\right) + \frac{1}{\rho_0}\nabla p' = 0\\ \frac{\partial p'}{\partial t} + \left(\vec{U}\cdot\nabla\right)p' + \gamma P(\nabla\cdot\vec{u}') + (\vec{u}'\cdot\nabla)P = -\frac{DP}{Dt}\end{gathered} \tag{3}$$

where ρ', $\vec{u}'$, p' are compressible perturbed flow density, velocity, pressure, γ is the ratio of the specific heats.

With this splitting method, the total velocity/pressure of the flow would be the sum of the incompressible flow velocity/pressure and acoustic velocity/pressure perturbation; hence the acoustic pressure obtained from incompressible fluid field also forms a loading on LVM surface

mesh. However, the calculation of the LPCE showed that the root-mean-value of the acoustic flow rate was only 0.9% of the incompressible flow rate, suggesting a weak coupling effect between the acoustic field and incompressible flow field. Therefore, the linear source-filter theory (*2*) was applied which assumes that the sound radiation in the tract is linearly coupled with the source generation in the syrinx so that the acoustic pressure does not affect vibrations. Because the dominant sound source of voice is a monopole source, the far-field acoustic pressure was calculated as below:

$$p' = \frac{\rho}{4\pi r}\frac{dQ}{dt} \tag{4}$$

where p' is the acoustic pressure, ρ is the air density, r is the distance from sound source to the microphone and dQ/dt stands for the time derivative of flow rate.

Power transfer from airflow to the vibration masses over a vibration cycle was calculated by multiplying the flow pressure force with the normal components of velocity vectors.

Parameterization of Vocal organ Geometry

The geometry of the airway and LVMs was reconstructed manually from the DiceCT scan of each specimen in Mimics (Materialise, Leuven, Belgium). We used every tenth slice of the original scan in this segmentation process. The surface mesh of the airway and LVMs was exported from Mimics. The LVM tetrahedral mesh for finite-element analysis was generated using ANSYS for P6, 15 and Gmsh for P11, 12, 13 and contained around 14,000 tetrahedral elements for each LVMs. Based on the anatomy, several places on the LVM surface were connected to syringeal solid structures, such as ossified tracheal rings, and therefore the motion at these places was constrained in the simulation. All other nodes were free to move and a traction boundary condition was applied.

Simulation Implementation and Conditions

The vocal organ mesh was immersed in a computational domain that was about 12×10×40 mm in dimensions and straight tubes were added to each bronchial inlet and tracheal outlet to avoid reverse flows. The domain was discretized with a non-uniform 64×48×128 Cartesian mesh with the highest grid density around the LVMs. A non-penetration non-slip boundary condition was applied at the airway wall. The part of the airway wall contacting the LVMs was the interface of the fluid-structure interactions on which the deformation and velocity of the LVM surface were transferred to the fluid solver and the fluid pressure on the LVM surface was transferred to the solid solver. The air density was set to 1.1455 kg/m^3 at 37°C. To reduce computational costs, we used a kinematic viscosity value four times higher than normal in air (is 6.6×10^{-5} m^2/s, at 37°C) as this reduces high frequency turbulence in the simulations (*34*).

We assumed the LVM in pigeons consists of isotropic material based on earlier histological sections (*55*). We assumed incompressible material properties and the Poisson ratio (v) was set to be 0.46 to avoid the singularity problem at v=0.5. The shear modulus (G) was obtained from the relationship of $G=E/2/(1+v)$ for isotropic materials. Tissue density was assumed to be 900 kg/m^3, equal to the density of fat.

For the dynamic simulations a time step of about 1.0 μs was utilized in the fluid and solid solvers. Two artificial non-slip and non-penetrable collision planes were 80 μm off the medial plane to enforce a finite 160 μm minimum gap between the two LVMs during closure. The LVM position was not allowed to exceed the corresponding collision plane during collision, to prevent failure of the solver due to the non-conserved mass in each zone. For each individual, the

simulation was performed on 16 processors and took about 1.5 days per vibration cycle. We obtained at least five cycles per individual.

495

Parameterization of LVM Elastic Moduli

To parameterize tissue elasticity for each individual, we developed a novel combined experimental/simulation approach. We used the finite-element LVM model to simulate LVM displacement as a function of transmural pressure as obtained in the experiment listed above for
500 a range of elastic moduli values. Uniform bronchial pressure steps were applied on the inner LVM surface and the maximum LVM gap was obtained for each pressure step. For each preparation, different elastic modulus values were randomly assigned to each static analysis solver resulting in a different displacement-transmural pressure slope. We determined the elastic modulus of the LVMs for each preparation in an iterative process by minimizing the difference
505 between the slopes of LVM displacement versus p_t measured in the experiment with the 3D mesh model.

EXP and SIM Data Analysis

Kinematic Analysis: To compare LVM kinematics we calculated the time-resolved glottal
510 opening (i.e. a glottovibrogram) from the left and right LVM position as measured in the highspeed video (EXP) and along the midline of the simulations (SIM). In this spatiotemporal matrix of glottal opening, we calculated the time-resolved minimum opening to detected values of complete closure, i.e., glottal opening = 0 µm for EXP and <100 µm for SIM over five cycles. The first instance of closure (defined as φ=0°) was used to calculate instantaneous vibration
515 period and fundamental frequency per cycle. The Open-Closed Quotient (OCQ) was defined as

the ratio of time open over time closed per cycle. The caudo-cranial (CC) wavespeed was calculated by linear regression of all data points of complete closure as a function of CC position and time in a cycle. For all kinematics parameters the mean over 5 cycles was used for further comparison.

To compare the LVM shape between EXP and SIM in more detail we first aligned EXP and SIM data over the midline and focused on the range where data was present for both EXP and SIM along the entire vibratory cycle (grey boxes in Fig 3). We calculated the LVM opening difference along the CC axis position between the mean shape of EXP and SIM over 5 consecutive cycles at phase φ=0, 60, 120, 180, 240 and 300°. To statistically compare LVM shapes we first calculated the cumulative difference between EXP and SIM opening along the CC axes (i.e. the area between LVM shapes). We then repeated this process 1000 times against a randomized EXP shape to create a distribution of mean area difference and tested whether the LVM shape and random shape were drawn from same continuous distribution using a two-sample Kolmogorov-Smirnov test per phase step.

Acoustical Analysis: To compare the acoustics of the model with the experimental obtained values we first resampled the SIM signal from about 1.2 MHz to 48 kHz using the resample function in Matlab. We low-pass filtered EXP and SIM sound signal at 20 kHz with a 3th order butterworth filter with zero-phase shift implementation (filtfilt.m function). Source level was defined at 1 meter from the source as SL = 20log(p/p_{ref}) + TL, with p the RMS pressure and p_{ref} =20 µPa, transmission loss TL = 20*log10(d) with distance d=12 cm. Spectral slope was calculated following (*6*). We first detected all peaks in a spectrum based of all 5 cycles. Spectral

slope was defined as the slope of the linear regression of amplitude over log10 frequency of all detected peaks multiplied by log10(2).

540

Statistical Analysis: One preparation was omitted for further analysis as fixation caused it to twist compared to the experimental situation. All data is presented as mean ± S.D. and comparisons were done with paired-sample t-tests at significance level p=0.05, unless mentioned otherwise.

545

FIGURES

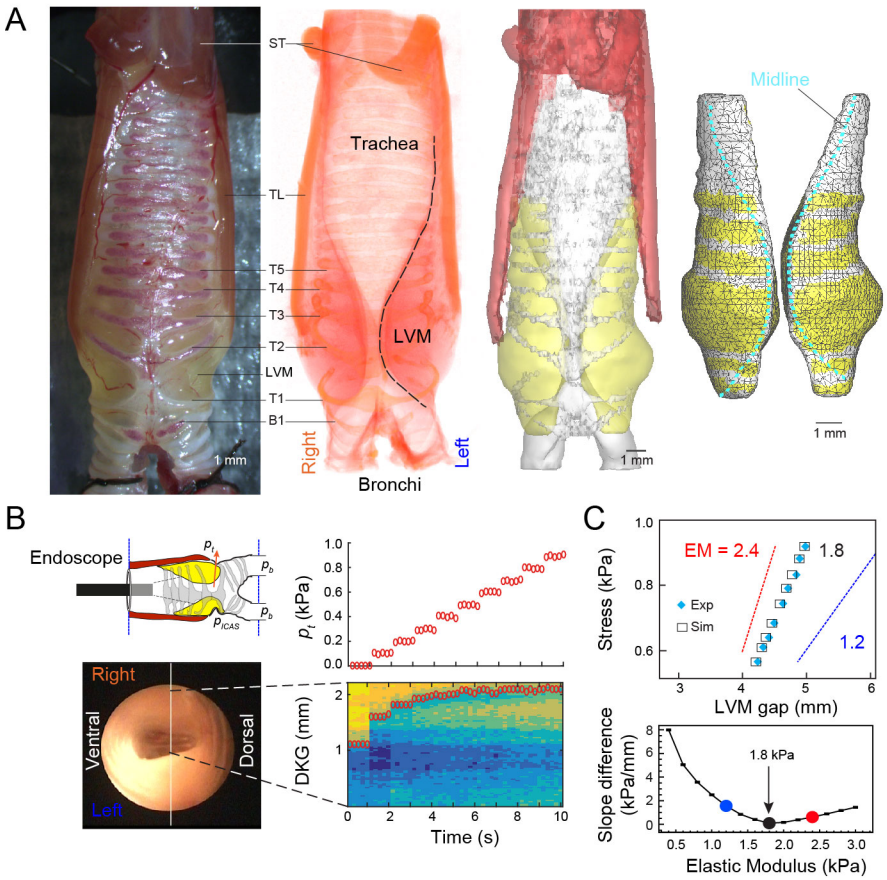
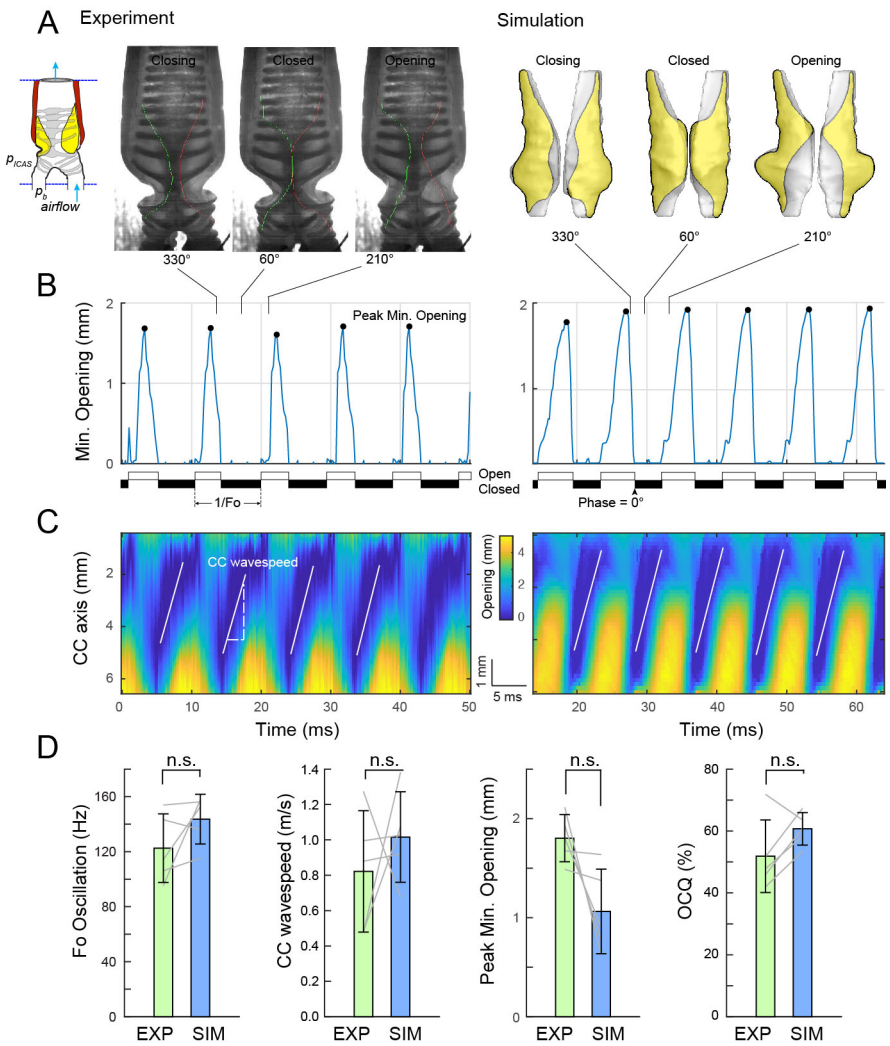


Figure 1. Parameterization of vocal organ geometry and tissue properties. (A) Workflow to parameterize FSAI model geometry (Pigeon P13), with from left to right: 1) photo of syrinx mounted in *in vitro* experimental setup, 2) voltex rendering of iodine contrasted µCT scan showing the bilateral Lateral Vibratory Masses (LVM), 3) 3D anatomy, and 4) Finite Element Mesh of LVM solid domain. Yellow and white mesh elements on LVM outer surface have free and zero displacement boundary conditions, respectively. (B) Tissue properties were determined by a combined experimental and modeling approach. A stepwise increase of (top) transmural pressure p_t caused a stepwise sideways displacement of the LVMs (bottom) as viewed by digital kymogram (DKG) along the white vertical line in the endoscopic image (left). **(C)** We used a reverse engineering approach to determine the LVM elastic modulus (EM) for each individual. Top: stress-gap width curve of experimental (blue diamonds) and modeled displacement for LVMs with EM=1.8 kPa (black squares). Also indicated are simulated gap widths of EM=1.2 (blue line) and 2.4 kPa (red line). Bottom: Minimal difference between experimental and simulated data indicates occurs at EM = 1.8 kPa for this individual (P6). B1, bronchial ring; T1,…, T5, tracheal rings; LVM, Lateral Vibratory Membrane; p_t, transmural pressure TL, tracheolateralis muscle; ST, sternotrachealis muscle.



570 **Figure 2. FSAI simulation accurately predicts key features of LVM kinematics.** (A) Trans-illuminated syrinx in experiment (EXP, left) and simulation (SIM, right) showing the LVM shape in coronal plane at 330, 60 and 210° of the vibratory cycle, where 0° is defined as the first frame of full closure. (B) The minimal glottal opening shows when the glottis in open and closed (white and black horizontal bars) and vibratory frequency *f*o. The open-closed quotient (OCQ) is the ratio between duration
575 of open and closed glottis. (C) Coronal glottovibrogram (cGVG) showing time-resolved glottal opening along caudocranial axis of 5 oscillations. The white lines indicate the CC wavespeed regression slopes based on the closed glottis. (D) Observed EXP data and SIM predictions are not significantly different for the above indicated four key kinematic parameters.

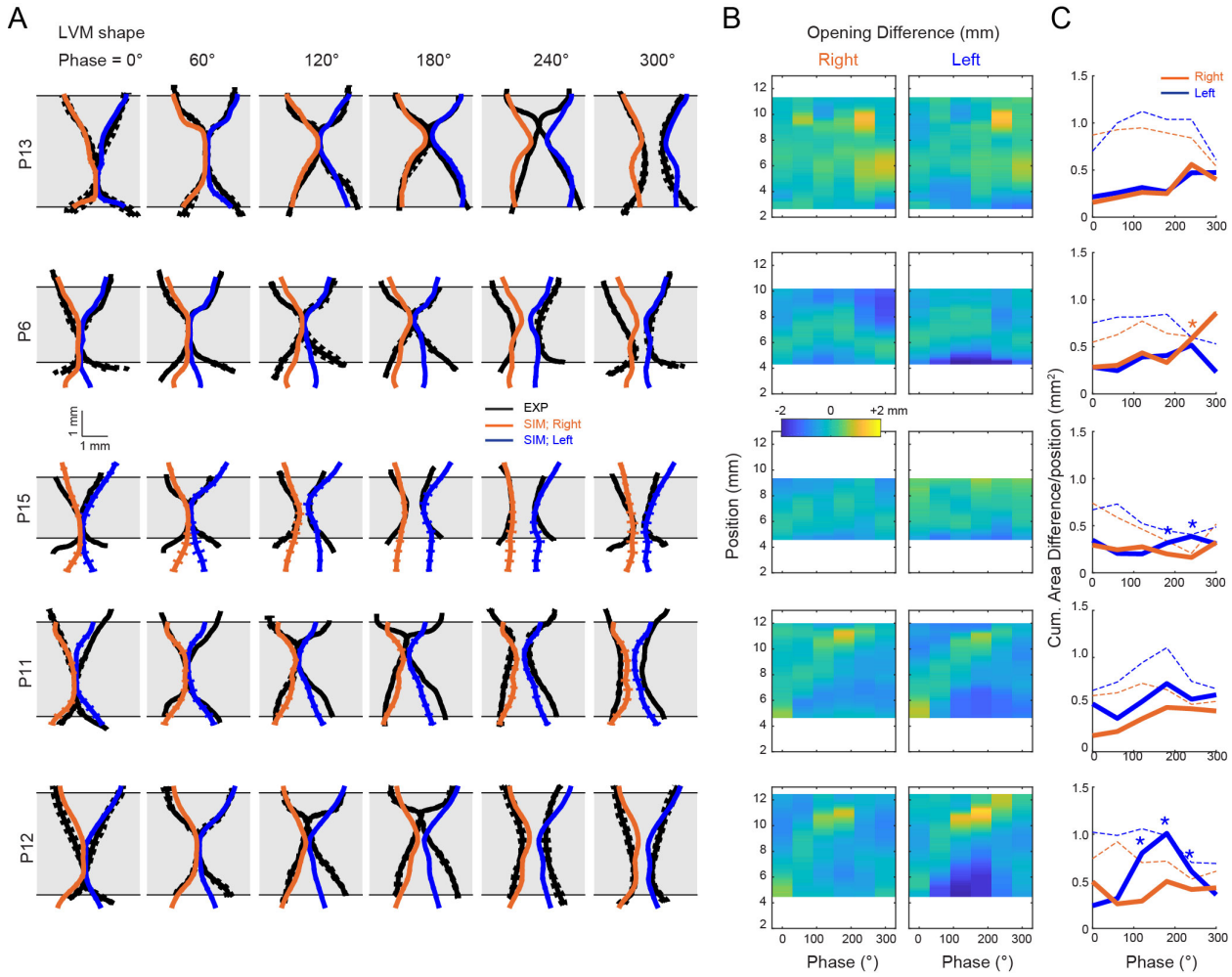


Figure 3. FSAI simulation accurately predicts within-cycle LVM shape. (A) Bilateral LVM shape of EXP and SIM in coronal plane at 0, 60,120, 180, 240, 300 degrees averaged over 5 cycles (mean (solid lines) ± S.D (dotted lines). (B) Lateral position difference between EXP and SIM left and right LVMs and (C) unsigned cumulative area difference between EXP and SIM shows that the predicted shapes are very similar to the observed shapes. The most pronounced discrepancies occur during late closed/early opening. Dotted lines are mean. Asterisks indicate significant difference ($p>0.01$) between SIM data and 1000x bootstrapped EXP shapes for 2-sample Kolmogorov-Smirnov tests.

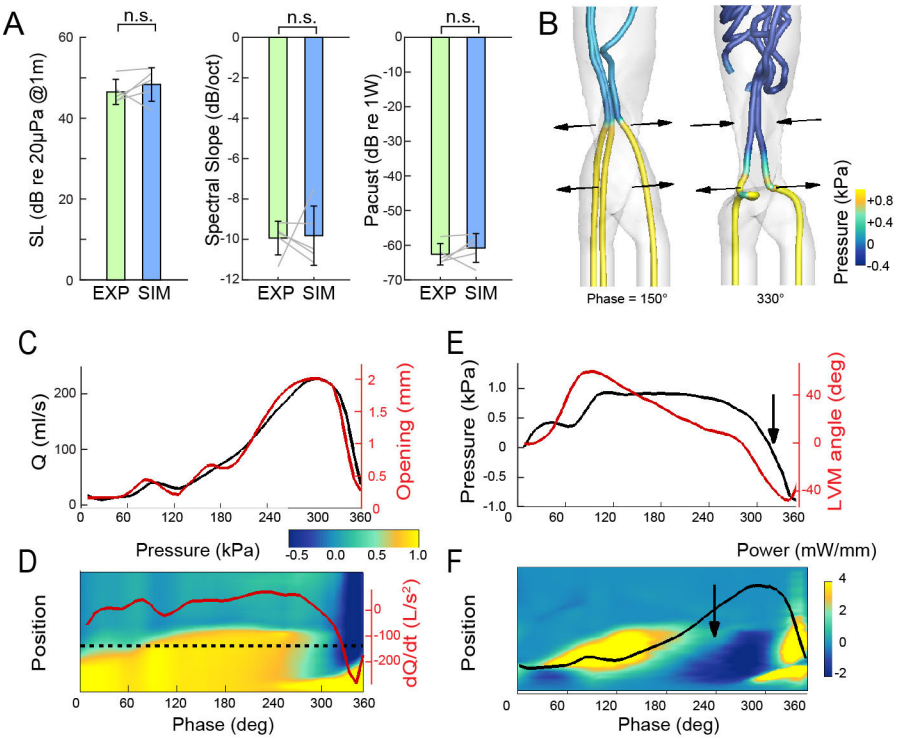
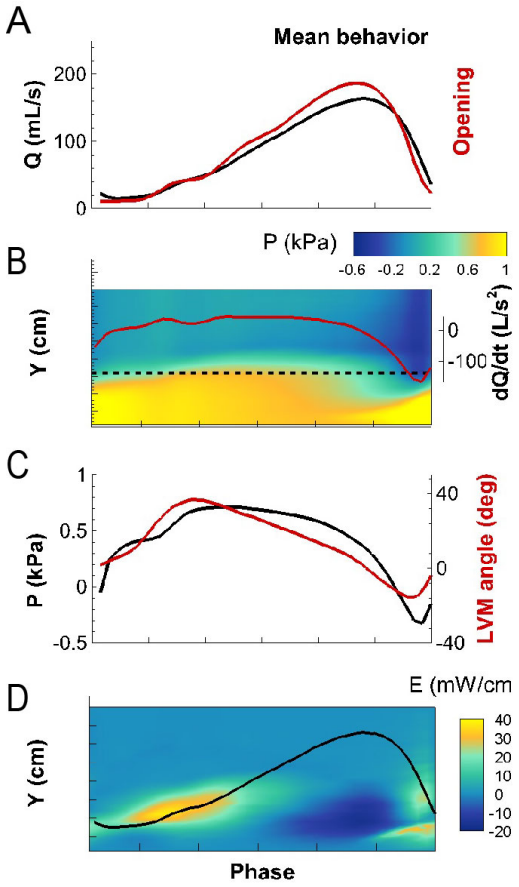


Figure 4. FSAI predictions of acoustics and spatiotemporal pressure and power profiles. (A) The SIM accurately predicted key acoustic parameters; source level (SL), spectral slope and acoustic power. (B) 3D flow inside the syrinx as indicated by flow streamlines (pressure color-contoured). Horizontal arrows indicate the motion direction of the LVMs. (C) Flow rate Q (black line) correlates strongly with glottal opening (red). (D) Spatiotemporal pressure distribution along the airway centerline over a cycle with superimposed dQ/dt (red solid line) as a proxy for sound pressure. (E) Glottal pressure (black line) evaluated at the horizontal dotted line in (D) with LVM angle (red line). (F) Spatiotemporal power transfer distribution from flow to LVM evaluated along the airway centerline with superimposed flow Q (black solid line).

SUPPLEMENTARY INFORMATION

Table S1. LVM finite element model statistics and elastic moduli per individual.

	Nodes		**Elements**		**EM**
ID	Left LVM	Right LVM	Left LVM	Right LVM	kPa
P6	1988	2278	8101	9838	1.8
P11	4253	4551	17062	18436	2.24
P12	4620	4693	18657	19131	2.4
P13	3988	3835	15737	15190	1.69
P15	2686	2159	11402	8532	4



615 **Figure S1. Mean FSAI prediction of spatiotemporal pressure, power profiles and acoustics.** Mean performance for all 5 individuals. (A) Flow rate Q (black line) correlates strongly with glottal opening (red). (B) Spatiotemporal pressure distribution along the airway centerline over a cycle with superimposed dQ/dt (red solid line) as a proxy for sound pressure. (C) Glottal pressure (black line) evaluated at the horizontal dotted line in (B) with LVM angle (red line). (D) Spatiotemporal power
620 transfer distribution from flow to LVM evaluated along the airway centerline with superimposed flow Q (black solid line).