Controlling chaotic vocal fold oscillations in the numerical production of vowel sounds

Oriol Guasch, Marc Freixes, Marc Arnela, Annemie van Hirtum

Controlling chaotic vocal fold oscillations in the numerical production of vowel sounds

Oriol Guasch [a,*], Marc Freixes [a], Marc Arnela [a], Annemie Van Hirtum [b]

[a] Human-Environment Research (HER) group, Department of Engineering, La Salle, Universitat Ramon Llull C/ Quatre Camins 30, 08022, Barcelona, Catalonia Spain
[b] University of Grenoble Alpes, CNRS, Grenoble INP, LEGI, 38000 Grenoble, France

Keywords:
Phonation pacemaker
Vocal fold mass model
Chaotic self-oscillations
Chaos control
Vowel production
Finite element method

Lumped mass models have been studied in depth to unveil the complex nonlinear physics of phonation. Even in the case of simple symmetric models, slight changes in muscle restoring forces or excessive subglottal pressure can cause abnormal or even chaotic vocal fold oscillations. In a recent work, it was shown that it was possible to device a theoretical pacemaker for phonation that could render the chaotic motion regular again. This consisted of attaching an additional mass–spring–damper system to the vocal fold model, the damping of which could be adjusted according to an altering energy chaos control strategy. The chaos of phonation is low-dimensional and one may wonder whether it has a profound effect in voice production and, if so, whether the proposed phonation pacemaker could compensate for it. For this purpose, we compute the time evolution of the glottal volume velocity generated by normal, chaotic and controlled oscillations of the vocal folds and convolve it with the impulse response of magnetic resonance imaging (MRI) geometries of the human vocal tract, corresponding to the vowels /ɑ/, /i/ and /u/. The impulse response for each vowel is obtained from the solution of the wave equation by the finite element method, when a Gaussian pulse is prescribed as a boundary condition in the glottis of the vocal tract. It will be demonstrated that the chaotic vibration of the vocal folds severely distorts the vowel sounds and that the proposed control strategy is able to recover with high quality the vowels produced in normal phonation. Audiovisual files are provided to support the objective results of the phenomena in terms of spectral and time analysis of the train of glottal pulses generated by the vocal folds and the produced vowel sounds.

1. Introduction

Phonation is the highly nonlinear process by which self-oscillations of the vocal folds result in the generation of a glottal volume velocity in the form of a train of glottal pulses. This acts as a source of acoustic waves that propagate through the vocal tract (VT) and are emitted outward from the mouth as a voiced sound. The fluid–structure interaction mechanism that triggers and maintains vocal fold (VF) self-oscillations is described, in general terms, by the myoelastic aerodynamic theory of phonation [1,2]. Since its formulation many decades ago, much work has been done to better understand intricacies of the physics that governs it. Research has involved from the development of low-dimensional lumped mass models [3–7] to complex computational models [8–12] and to experiments on mechanical replicas [13–16].

Healthy phonation is characterized by regular self-oscillations of the VFs, which produce a glottal volume velocity that is driven by a fundamental frequency, its harmonics and some randomness noise. This signal is often said to be quasi-periodic, not in the sense of a dynamical system governed by incommensurable frequencies, but by being sufficiently regular in a loose sense. In most voice pathologies this regular motion of the VFs becomes disturbed. For example, unilateral paralysis and lesions due to nodules, polyps or cysts induce an asymmetry in the VFS that may cause them to vibrate chaotically rather than quasi-periodically. In Parkinson's disease and in paresis the VFs can remain symmetric but their physical parameters (e.g., stiffness) change in such a way that the VFs motion can become also chaotic instead of regular.

Studying the physics of pathological voice is very challenging. As for computational models, and despite some attempts [17,18], they are very costly and there is still a long way to go even to perform complete simulations of healthy phonation [12]. Attempts have also been made to reproduce voice disorders with mechanical replicas for

* Corresponding author.
E-mail address: oriol.guasch@salle.url.edu (O. Guasch).

left and right VF asymmetries that represent, for example, a unilateral paralysis [19,20]. However, it was not until very recently that synthetic VFs with inclusions emulating a polyp have been characterized [21,22] and their chaotic motion experimentally measured [23]. In view of the difficulties and cost of performing complex fluid–structure interaction simulations and delicate mechanical replica experiments, lumped mass models of the VFs remain a powerful tool for investigating the physics of unhealthy phonation and possible remedies. Probably, the most celebrated VF models are the two-mass model of Ishizaka and Flanagan [3] and the cover-layer model (three masses) of Story and Titze [5]. Both have been used to understand the transition from quasi-periodic to chaotic motion for VFs with asymmetric parameter values [24], symmetric VFs with abnormal stiffness values to investigate aspects of Parkinson's disease like vocal tremor [25,26] and VFs with polyps [27,28]. Unfortunately, although the physics of voice disorders are fairly well explained, there are not many options to remedy them.

In a recent work [29], a symmetric two-mass model was used to explore the idea of designing a VF pacemaker to regulate the aperiodic and/or chaotic motion of VFs. The pacemaker consisted of a patch of an ideal smart material glued to the VFs whose parameters could be dynamically adjusted as the VF vibrated. In the lumped mass-model, it was represented by an additional mass attached to the VFs whose damping or stiffness could be rapidly modified according to a chaos control strategy. In particular, the altering energy control method was used [30,31], which previously proved to be very useful for the classical Van der Pol and Duffing non-linear oscillators, as well as for problems involving collisions [32]. The results were very encouraging in the sense that regular phonation could be restored, but no link was established with respect to the efficiency of the pacemaker in relation to the generation of voiced sounds. In fact, this is the main purpose of the pacemaker and therefore the aim of this work.

Our goal and contribution in this paper is to analyze whether the theoretical pacemaker in [29] could have a significant impact on voice generation. For this purpose, we produce numerical vowels /ɑ/, /i/ and /u/ with the finite element method (FEM) using regular, chaotic and pacemaker-controlled glottal volume velocities obtained from a lumped mass model of the VFs. It is worth noting that many voice disorders can be detected by analyzing the acoustics of sustained vowels sounds [33,34]. To generate the train of glottal pulses we consider a slightly improved model over that of [29], which takes into account cubic muscle restoring forces for the VFs, the mass and stiffness of the pacemaker and the option to include or not vibrato. Vibrato is a voice/musical effect consisting in a pulsating change of the pitch. Here, it is simulated by introducing time variations in the stiffness of the lower masses of the VF model, which poses more challenges to the chaos controlling mechanism. More physical-based options for simulating vibrato can be found in [35,36]. The cases of regular (healthy) phonation and chaotic (unhealthy) phonation, due to an excessive value of the coupling stiffness between the upper and lower masses of the VFs are investigated, with and without vibrato. The chaos control strategy is applied to the unhealthy situations and its performance is inspected. Bifurcation and phase state plots for the motion of the VF masses are provided. Special emphasis is then placed on our variable of interest, the glottal volume velocity, for which time-delay coordinate plots and spectra are presented. The glottal volume velocity is used as the acoustic source term to produce vowel sounds. Realistic three-dimensional magnetic resonance imaging (MRI) geometries of the vocal tracts (VTs) corresponding to vowels /ɑ/, /i/ and /u/ [37] are considered and their impulse response is computed solving the wave equation for a Gaussian pulse excitation at the glottis, using an in-house FEM code as in [38,39] (see also [40]). The regular, chaotic and controlled glottal volume velocities, including or not vibrato, are convolved with the VT impulse response of each vowel to obtain the corresponding acoustic pressure. The acoustic power spectrum, time evolution and spectrogram for /ɑ/, /i/ and /u/ are provided to show the potential of the theoretical pacemaker to improve the vowel sounds generated by unhealthy phonation. Supplementary material consisting of audiovisual files to listen to the generated sounds is also made available.

The paper is organized as follows. Section 2 deals with theory and modeling. First, it presents the non-linear two mass model of the VFs and details its particularities (Section 2.1). The implemented chaos control strategy is described in Section 2.2, while the procedure for computing the VT impulse response and the acoustic pressure for each vowel is provided in Section 2.3. Section 3 is devoted to numerical simulations and discussion of the results. Section 3.1 presents the numerical strategy to solve all the involved equations, while Section 3.2 analyzes regular, chaotic and controlled phonation. Vowel production is addressed in Section 3.2. Conclusions close the article in Section 4.

2. Modeling of the vocal folds and vocal tract to generate vowels

2.1. Nonlinear two-mass model of the vocal folds

In this section, we present the two-mass models to be used in our simulations of VF vibrations. As seen in Fig. 1, three cases are considered. The first one (Fig. 1a) corresponds to healthy phonation and is basically described by the model in [24,25] with some modifications detailed below. The left VFs are represented by an upper and a lower mass, with values m_{1l} and m_{2l}, which are connected to rigid laryngeal walls by dampers with damping values r_{il}, $i = 1,2$, and springs of stiffness k_{il}, $i = 1,2$. The dampers r_{il} account for the mass stickiness when the surfaces of the left and right vocal folds contact, while k_{il} determines the tension of the vocal folds mostly regulated by the contraction of the anterior cricothyroid muscle. The relative motion between the upper and lower masses determines the mucosal wave. This relative motion is limited by the flexural stiffness in the lateral direction of the VFs, characterized by k_{cl}. More details on the physiological meaning of the model parameters are provided in [3,24,25]. The two-mass model system is driven by the glottal airflow emanating from the lungs, which obeys the Bernoulli equation. When the VFs are closed, the subglottal pressure increases until it reaches a threshold value that opens them. Then, the pressure drops and the muscle restoring forces close the VFs. The subglottal pressure increases again and the process repeats, resulting in VF self-oscillations and the establishment of regular phonation.

In the second case (Fig. 1b), we simply modify the stiffness k_{cl} of the spring connecting the two masses to a new value that transitions the motion of the VF from regular to chaotic, beyond a given subglottal pressure [25]. Finally, in the third case (Fig. 1c) we attach a phonation pacemaker to the lower VF mass [29]. The pacemaker consists of a small mass–spring–damper resonator with values m_{sm}, k_{sm} and r_{sm} and is assumed to be made of an ideal smart material such that r_{sm} could be tuned on demand to make the VFs motion regular again. It should be noted that the two-mass model in this paper has some variations from the previous one in [29]. This time we do not neglect the values of the mass, m_{sm}, and stiffness, k_{sm}, of the pacemaker versus m_{1l} and k_{1l}. In addition, we assume a cubic restoring force for mass springs, k_{1l} and k_{2l}, which better represents the behavior of the vocal fold tissue, as done in [3]. Vibrato [36] is approximated in the model by letting k_{1l} vary with time. This introduces a slight oscillation in the fundamental frequency of the train of glottal pulses, which helps to generate more natural vowel sounds.

Let us now present the ODE system governing the dynamics of the two-mass model of Fig. 1c (those of Fig. 1a and b are easily obtained by introducing some simplifications to it, see below). Following [3,24,25,29], we apply Newton's second law to the upper and lower masses of the VFs and treating their velocities, $v_{1l} = \dot{x}_{1l}$ and $v_{2l} = \dot{x}_{2l}$, as

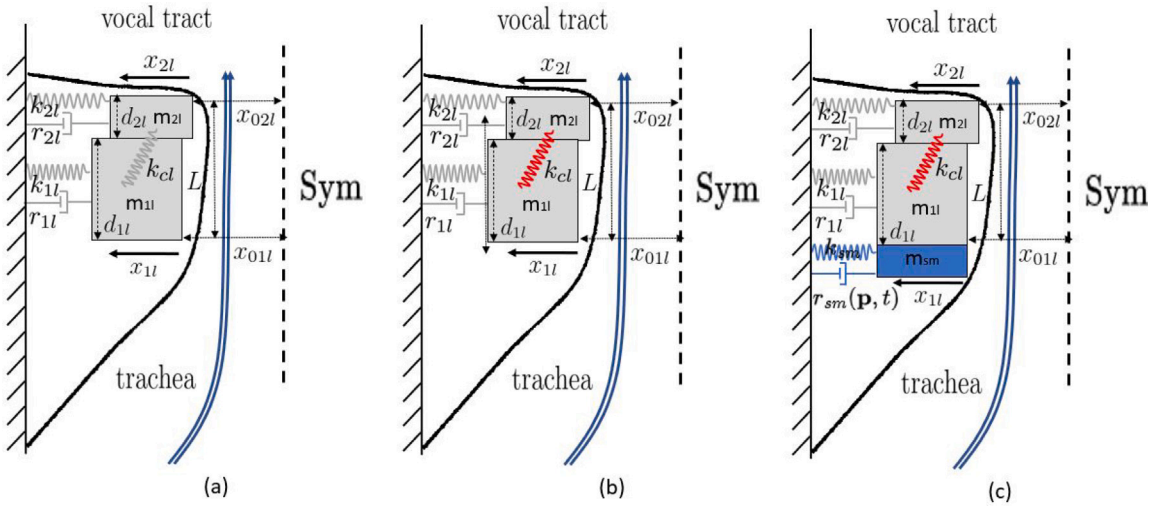

Fig. 1. Sketch of a symmetric two-mass model of the vocal folds (VFs) driven by the glottal airflow from the trachea to the vocal tract. (a) All model parameters have appropriate values leading to healthy, regular phonation. (b) The contact stiffness between the masses, k_{cl} (red color), is higher than normal, resulting in chaotic VF oscillations and abnormal phonation beyond a given subglottal pressure. (c) A phonation pacemaker consisting of a mass–spring–damper resonator (blue color) is attached to the lower mass and its damping is controlled to transition from chaotic VF vibrations to regular ones. The dynamics of these models are described by Eq. (1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

independent variables, it can be shown that the dynamics of the current VF model is governed by the non-linear matrix system,

$$\begin{pmatrix} \dot{\boldsymbol{v}} \\ \dot{\boldsymbol{x}} \end{pmatrix} = \begin{pmatrix} -\mathbf{M}^{-1}\mathbf{C}(t,\boldsymbol{p}) & -\mathbf{M}^{-1}(\mathbf{K}(t,\boldsymbol{x})+\boldsymbol{\Theta}(\boldsymbol{x})) \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{x} \end{pmatrix} + \begin{pmatrix} \mathbf{M}^{-1}\boldsymbol{f}_v(\boldsymbol{x}) \\ \mathbf{0} \end{pmatrix}, \tag{1}$$

with $\boldsymbol{x} = (x_{1l}, x_{2l})^\top$ and $\boldsymbol{v} = (v_{1l}, v_{2l})^\top$ respectively standing for the mass displacement and velocity vectors. Note that the pacemaker mass is glued to m_{1l} so it will have its displacement and velocity. The force vector $\boldsymbol{f}_v$ in Eq. (1) has the expression $\boldsymbol{f}_v = (P_1 L d_{1l}, 0)^\top$, where L is the glottis length, d_{1l} the thickness of m_{1l} and P_1 the glottal pressure,

$$P_1 = P_s \left[1 - \Theta(a_{\min}) \left(\frac{a_{\min}}{a_1}\right)^2\right] \Theta(a_1). \tag{2}$$

In Eq. (2), P_s denotes the subglottal pressure and $a_{\min} = a_1$ for $x_{1l} < x_{2l}$ while $a_{\min} = a_2$ for $x_{2l} \le x_{1l}$. Variables $a_i = a_{0i} + 2Lx_{il}$ $(i = 1, 2)$ are the lower and upper glottal areas with $a_{0i} = 2Lx_{0il}$ $(i = 1, 2)$ being the lower and upper glottal rest areas. x_{0il} designates the distance from m_{il} to the midline in the rest position. $\Theta(z)$ is the collision function given by $\Theta(z) = \tanh(50z/z_0)$ if $z > 0$ and $\Theta(z) = 0$ f $z \le 0$. z_0 is taken as a_{01} in the calculations.

As for the block matrix of Eq. (1), its second row contains the identity matrix $\mathbf{I}$ and the zero matrix $\mathbf{0}$, while in the first row we identify the mass matrix $\mathbf{M}$, the damping matrix $\mathbf{C}$, the stiffness matrix $\mathbf{K}$ and the collision matrix $\boldsymbol{\Theta}$. Note that the damping matrix explicitly depends on time and on a vector of variables and parameters $\boldsymbol{p}$ that will be determined by the chaos control strategy. The stiffness matrix $\mathbf{K}$ also exhibits explicit dependence on time, as well as on the displacement vector $\boldsymbol{x}$. The collision matrix also varies with $\boldsymbol{x}$. The detailed expressions of these matrices are,

$$\mathbf{M} = \mathbf{M}_{\mathrm{VF}} + \mathbf{M}_{\mathrm{P}} = \mathrm{diag}(m_{1l}, m_{2l}) + \mathrm{diag}(m_{sm}, 0), \tag{3a}$$

$$\mathbf{C} = \mathbf{C}_{\mathrm{VF}} + \mathbf{C}_{\mathrm{P}} = \mathrm{diag}(r_{1l}, r_{2l}) + \mathrm{diag}(r_{sm}(t,\boldsymbol{p}), 0), \tag{3b}$$

$$\mathbf{K} = \mathbf{K}_{\mathrm{VF}} + \mathbf{K}_{\mathrm{P}} = \begin{pmatrix} \tilde{k}_{1l} + k_{cl} & -k_{cl} \\ -k_{cl} & \tilde{k}_{2l} + k_{cl} \end{pmatrix} + \begin{pmatrix} k_{sm} & 0 \\ 0 & 0 \end{pmatrix}, \tag{3c}$$

$$\boldsymbol{\Theta} = \mathrm{diag}\left(\Theta(-a_1)c_{1l}\left[\frac{a_{01}}{2Lx_{1l}} + 1\right], \Theta(-a_2)c_{2l}\left[\frac{a_{02}}{2Lx_{2l}} + 1\right]\right). \tag{3d}$$

As seen, we have split the mass, damping and stiffness matrices into a component arising from the standard two-mass VF model (matrices with subscript VF) and a component coming from the phonation pacemaker (subscript P). As mentioned above, the mass and stiffness of the pacemaker, m_{sm} and k_{sm}, are now taken into account in the model. However, they will play a passive role in this work because the control of chaos will be exerted by the damping term $r_{sm}(\boldsymbol{p}, t)$ in $\mathbf{C}_{\mathrm{P}}$, as will be explained in next subsection. Regarding the VF stiffness matrix, $\mathbf{K}_{\mathrm{VF}}$, its lower mass element is $\tilde{k}_{1l} = k_{1l} T_{k_{1l}}(t) X_{k_{1l}}(x_{1l})$. The time dependent function,

$$T_{k_{1l}}(t) = 10^{A_v \sin(\omega_v t/1000)} \tag{4}$$

allows us to approximately simulate a vibrato with amplitude A_v and modulating frequency f_v ($\omega_v = 2\pi f_v$). Note that we do not claim here that this modification of $\tilde{k}_{1l}$ induces a vibrato based on realistic physical grounds, as in [35], but it leads to similar effects, as will be shown in the spectrograms of Section 3.3.2 (see [36]) and perceived in the audios of the supplementary material. On the other hand, the function depending on space

$$X_{k_{1l}}(x_{1l}) = (1 + \eta_k x_{1l}^2) \tag{5}$$

accounts for the nonlinearity of the muscle restoring force, characterized by the coefficient η_k. As for the upper mass spring, we take $\tilde{k}_{2l} = k_{2l}(1 + \eta_k x_{2l}^2)$. Lastly, the values c_{il}, $i = 1, 2$, in $\boldsymbol{\Theta}$ are additional spring constants during collision.

As observed in Eqs. (1)–(3), nonlinearity enters the two-mass model through matrices $\mathbf{K}$, $\boldsymbol{\Theta}$, and the external vector $\boldsymbol{f}_v$. The solution to Eq. (1) gives the displacements and velocities of the masses in the chaos-controlled system of Fig. 1c. The parameters to be used for the numerical simulations in Section 3 are presented in Table 1. To obtain the solution for the chaotic and regular systems in Fig. 1b and a, respectively, we simply need to set $m_{sm} = k_{sm} = r_{sm}(\boldsymbol{p}, t) = 0$ in Eqs. (1)–(3). The parameters in Table 1 then become those in [3,24,25] (if vibrato is also removed, i.e., $A_v = 0$). Note that for the chaotic motion of the system in Fig. 1b we need to take $k_{cl} = 0.09$ g ms^{-2}, while for regular oscillations of the system in Fig. 1a we set $k_{cl} = 0.025$ g ms^{-2} (see Table 1 and [25]).

To conclude this section, let us point out that the solution of the system Eq. (1) will provide us with the displacements and velocities of the upper and lower masses of the VF model. However, for the generation of vowel sounds, our variable of maximum interest is the glottal flux that is characterized by its volume velocity,

$$U_a = \left(\frac{2P_s}{\rho_0}\right)^{1/2} a_{\min} \Theta(a_{\min}), \tag{6}$$

Table 1
Physical and geometrical parameters of the VFs (see [3,24,25] for the models without chaos control).

Physical parameters	Geometrical parameters
$m_{1l} = 0.125$ g	$d_{1l} = 0.25$ cm
$m_{2l} = 0.025$ g	$d_{2l} = 0.05$ cm
$r_{1l} = 0.02$ g ms^{-1}	$L = 1.4$ cm
$r_{2l} = 0.01$ g ms^{-1}	$a_{01} = 0.05$ cm^2
$k_{1l} = 0.08$ g ms^{-2}	$a_{02} = 0.05$ cm^2
$k_{2l} = 0.008$ g ms^{-2}	
$k_{cl} = 0.025$ g ms^{-2} (regular)	
$k_{cl} = 0.09$ g ms^{-2} (chaos)	
$c_{1l} = 3k_{1l}$ g ms^{-2}	
$c_{2l} = 3k_{2l}$ g ms^{-2}	
$m_{sm} = 0.0125$ g	
$k_{sm} = 0.008$ g ms^{-2}	
$r^0_{sm} = 0.002$ g ms^{-1}	
$\eta_k = 0.15$ cm^{-2}	
$\rho_0 = 0.00114$ g cm^{-3}	
$c_0 = 35$ cm ms^{-1}	
$A_v = 0.03$	
$f_v = 5$ Hz	

which can be calculated from the solution of the ODE system at each time step. ρ_0 is the air density and the subscript $a = \text{reg}, \text{cha}, \text{con}$ is used in the following to indicate the regime of the glottal volume velocity. U_a becomes the source of the acoustic waves in the vocal tract and in Section 2.3 below we will show how to generate vowels using it.

2.2. Chaos control strategy using time-varying damping in the phonation pacemaker

The chaos control strategy for the VF phonation pacemaker is based on the altering energy method proposed in [30–32]. In short, the idea behind this method is to apply an external control force, $\boldsymbol{f}_c(t, \boldsymbol{p})$, to the system (e.g., a set of nonlinear oscillators), such that the power $\boldsymbol{v}^\top \boldsymbol{f}_c(t, \boldsymbol{p})$ is strictly positive or negative. This changes the mechanical energy of the system to a new value where its motion is no longer chaotic, but regular. As explained in [29], an external force could hardly be applied to the VFs in practice, so it was proposed to replace it by an internal one, exerted by an ideal smart material (blue resonator in Fig. 1c) whose parameters can be modified as required. The altering energy approach for the current two-mass model of the VFs can be understood as follows.

The mechanical energy of the VF model is given by

$$E = \frac{1}{2}\boldsymbol{v}^\top \mathbf{M}\boldsymbol{v} + \frac{1}{2}\boldsymbol{x}^\top \mathbf{K}(t, \boldsymbol{x})\boldsymbol{x}, \tag{7}$$

and its derivative with respect to time by,

$$\dot{E} = \boldsymbol{v}^\top (\mathbf{M}\dot{\boldsymbol{v}} + \mathbf{K}\boldsymbol{x}) + \frac{1}{2}\boldsymbol{v}^\top \frac{d\mathbf{K}}{d\boldsymbol{x}} |\boldsymbol{x}|^2 + \frac{1}{2}\boldsymbol{x}^\top \dot{\mathbf{K}}\boldsymbol{x} \tag{8}$$

where we have used that $\mathbf{K}$ is symmetric and that it depends explicitly on time and the displacement vector. Using the first row of Eq. (1) to rearrange the first term in the l.h.s of Eq. (8), and taking into account that all block matrices of Eq. (3) are symmetric and the split of the damping matrix into $\mathbf{C} = \mathbf{C}_{\text{VF}} + \mathbf{C}_{\text{P}}$, we obtain

$$\dot{E} = \boldsymbol{v}^\top \left[-\mathbf{C}_{\text{VF}}\boldsymbol{v} - \boldsymbol{\Theta}(\boldsymbol{x})\boldsymbol{x} + \boldsymbol{f}_{\boldsymbol{v}}(\boldsymbol{x}) + \frac{1}{2}\frac{d\mathbf{K}}{d\boldsymbol{x}} |\boldsymbol{x}|^2 + \mathbf{C}_{\text{P}}(t, \boldsymbol{p})\boldsymbol{v} \right] + \frac{1}{2}\boldsymbol{x}^\top \dot{\mathbf{K}}\boldsymbol{x}. \tag{9}$$

Here, the internal power injected or removed by the pacemaker, $\boldsymbol{v}^\top \mathbf{C}_{\text{P}}(t, \boldsymbol{p})\boldsymbol{v}$ plays the role of the external control power $\boldsymbol{v}^\top \boldsymbol{f}_c(t, \boldsymbol{p})$ in the standard altering energy method of chaos control. Comparing Eq. (9) with Eq. (9) in [29] new terms can be identified because of the dependence of the stiffness matrix on t and $\boldsymbol{x}$ in the current model. Nonetheless, we will see that the same rules for damping control introduced in [29] also work for this case. In particular, several strategies were followed in [29]: damping control by attaching the pacemaker to the upper or lower VF masses, stiffness control by modifying k_{sm} and even a feedback time damping control strategy based on volume velocity to cope with non-constant subglottal pressure situations. The later need not be considered for vowel generation, and of all the control options tested, the most reasonable one turns out to be to glue the pacemaker resonator to the lower mass and dynamically adjust its damping.

Therefore, our choice in this paper is simply to take $r_{sm}(\boldsymbol{p}, t)$ in $\mathbf{C}_{\text{P}}$ to be

$$r_{sm}(\boldsymbol{p}, t) = r^0_{sm}\,\text{sgn}(v_{1l}) \tag{10}$$

where sgn stands for the sign operator and $\left|r^0_{sm}\right| \ll |r_{1l}|$. Then, the variables and parameters in $\boldsymbol{p}$ are simply the amplitude constant r^0_{sm} and the velocity of the lower mass v_{1l}.

Before going any further, a few words should be said about the possibility of manufacturing a real phonation pacemaker. Although the aim of the original work in [29] and the current one is to explore the potential of such a device using low-dimensional theoretical models of phonation, the idea of designing a biocompatible smart material that meets the requirements needed for a real phonation pacemaker is, in principle, feasible. The pacemaker would consist of a patch of a smart material plus a sensor to capture the vocal fold oscillations, a processor to calculate the signal to be applied to the material and an actuator to do so. The stimulation of the material will change its mechanical properties (e.g., its viscoelasticity). The stimulus could be either stress (typically up to 10 kPa during vocal cord collision), deformation (typically on the order of a millimeter) or an electrical pulse. Most likely, surgery would be required to implant the pacemaker into the damaged vocal cord tissue. As for the responsive material, different options could be considered. Among them, reasonable ones are those based on engineering electroactivated thermoplastic biomedical silicone elastomers (EA-TPE) [41], dielectric silicone elastomers (DE) [42–45], combined or not with yield stress fluids [46–50] due to their promising properties for biomedical applications, such as remarkable electro-actuation strains with high electromechanical efficiencies and low strain-cycling hysteresis across a broad range of electric fields. In addition, recent silicone polymer 3D prototyping [51,52] could be used to print customized patches of smart material for patients. However, there is still much preliminary theoretical and experimental work to be done before tackling the fabrication of an actual phonation pacemaker.

2.3. Vocal tract characterization and vowel generation

Having solved Eq. (1) and computed the time evolution of the glottal volume velocity from Eq. (6) for the cases of regular, chaotic and controlled motion of the VFs, namely $U_{\text{reg}}(t)$, $U_{\text{cha}}(t)$ and $U_{\text{con}}(t)$, there exist different options to numerically produce a vowel sound in the framework of the FEM. $U_a(t)$, with $a = \text{reg}, \text{cha}, \text{con}$ is the source of acoustic waves inside the VT, so given a VT geometry, the standard procedure is to directly prescribe $U_a(t)$ as a boundary condition on the surface of the glottis, and solve the wave equation with the FEM inside the VT and in a region outside the mouth to allow acoustic waves to propagate outwards. By collecting the acoustic pressure, $p(t)$, at a point near the mouth exit and converting it into an audio file, the vowel sound corresponding to that VT geometry is obtained. This procedure was the option followed in [53] to generate vowels and in [54] to produce diphthongs.

Unfortunately, the computational cost of three-dimensional finite element simulations is very high for large computational domains. Since a linear wave equation is normally used in VT acoustics, if several simulations are to be performed on the same VT geometry it is best to compute the VT impulse response, $h(t)$, and then convolve it with $U_a(t)$ to obtain the vowel sound for each phonation regime. The vowel sound is then given by (see e.g. [55]),

$$p(t) = U_a(t) * h(t) = \int h(t - \tau)U_a(\tau)\mathrm{d}\tau. \tag{11}$$

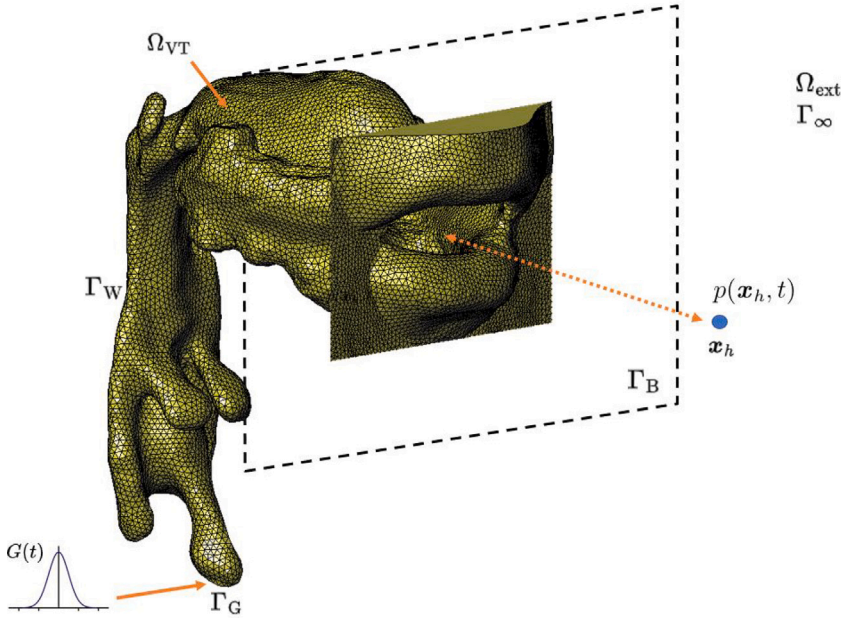

Fig. 2. Computational domain and scheme for calculating the impulse response of the VT, $h(t)$. A Gaussian pulse $G(t)$ is prescribed at the glottis Γ_{G} and the computed acoustic pressure is collected at a point $\boldsymbol{x}_h$ in front of the mouth (blue dot). The exterior domain Ω_∞ and its boundary Γ_∞ are not represented for clarity. Also, only a portion of the rigid baffle Γ_{B} is shown. Ω_{VT} stands for inner volume of the vocal tract and Γ_{W} for its walls. This particular VT corresponds to that of vowel /ɑ/.

Alternatively, one can make use of the convolution theorem to work in the frequency domain. For instance, this was the option chosen in [39] to evaluate the glottal source contributions to high-frequencies for different refinements of the VT geometry and also the option chosen in [56] to produce vowel–vowel utterances. It is also the choice made in this paper. Given a generic function $g(t)$, let us use the hat symbol to denote its Fourier transform, i.e., $\hat{g}(f) \equiv \mathcal{F}(g)$. The acoustic pressure at the mouth exit for a given vowel can then be calculated as,

$$p(t) = \mathcal{F}^{-1}[\hat{U}_a(f)\hat{h}(f)]. \tag{12}$$

$\hat{U}_a(f)$ is obtained directly from the Fourier transform of $U_a(t)$. As for $\hat{h}(f)$, it is the Fourier transform of the impulse response of the VT, $h(t)$, and is referred to as the VT transfer function (VTTF).

In this work, we compute the VT impulse response from the solution of the problem depicted in Fig. 2. We begin by computing the time evolution of the acoustic pressure, $p(\boldsymbol{x}_h, t)$, at a point $\boldsymbol{x}_h$ in front of the mouth when a Gaussian impulse is input at the glottis. Let us formulate the mathematical problem in some detail and denote the computational domain in the figure by $\Omega = \Omega_{\text{VT}} \cup \Omega_{\text{ext}}$ with boundary $\partial\Omega = \Gamma_{\text{G}} \cup \Gamma_{\text{W}} \cup \Gamma_{\text{B}} \cup \Gamma_\infty$, such that $\Gamma_{\text{G}} \cap \Gamma_{\text{W}} \cap \Gamma_{\text{B}} \cap \Gamma_\infty = \emptyset$. Here Ω_{VT} represents the volume inside the VT and Ω_{ext} the exterior one. Γ_{G} is the glottis surface and Γ_{W} the VT walls. Γ_{B} is a rigid baffle where the mouth is inserted. Our objective is to compute the acoustic pressure field $p(\boldsymbol{x}, t)$ in Ω that satisfies the wave equation with initial and boundary conditions,

$$-\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \nabla^2 p = 0 \quad \text{in } \Omega,\ t > 0, \tag{13a}$$

$$p = 0, \frac{\partial p}{\partial t} = 0 \quad \text{in } \Omega,\ t = 0, \tag{13b}$$

$$\nabla p \cdot \boldsymbol{n} = -\frac{\rho_0}{S_{\text{G}}}\frac{\partial G}{\partial t} \quad \text{on } \Gamma_{\text{G}},\ t > 0, \tag{13c}$$

$$\nabla p \cdot \boldsymbol{n} = -\frac{\mu}{c_0}\frac{\partial p}{\partial t} \quad \text{on } \Gamma_{\text{W}},\ t > 0, \tag{13d}$$

$$\nabla p \cdot \boldsymbol{n} = 0 \quad \text{on } \Gamma_{\text{B}},\ t > 0, \tag{13e}$$

$$\nabla p \cdot \boldsymbol{n} = \frac{1}{c_0}\frac{\partial p}{\partial t} \quad \text{on } \Gamma_\infty,\ t > 0, \tag{13f}$$

In Eq. (13a), c_0 represents the speed of sound. Eq. (13b) stands for the initial conditions of the problem. Regarding the boundary conditions, we impose a Gaussian pulse volume velocity, $G(t)$, on the glottis Γ_{G} (S_{G} denoting its surface, see Eq. (13c)), given by

$$G(t) = e^{-[(t-T_G)/0.29T_G]^2} \tag{14}$$

with $T_G = 0.646/f_c$, $f_c = 10$ kHz and $[G(t)] = \text{m}^3/\text{s}$, [38–40]. The coefficient μ for the VT wall losses in Eq. (13d) is related to the wall impedance, Z_{W}, by $\mu = Z_0/Z_{\text{W}}$ with $Z_0 = \rho_0 c_0$ and $Z_{\text{W}} = 83\,666\ \text{kg/m}^2$ s, see [57]. Eq. (13e) simply states that the baffle at the mouth exit is rigid. Finally, the Sommerfeld non-reflection condition (Eq. (13f)) is imposed on the boundary of the outer part of the computational domain, Γ_∞. In practice, this condition is not enough to guarantee wave-free reflection from Γ_∞ and is replaced by a perfectly matched layer (PML), which adds supplementary variables and equations to Eq. (13a), see [58]. The problem of Eq. (13a) with initial conditions in (13b) and boundary conditions in Eqs. (13c)–(13f) has been solved in this work using the numerical strategy described in the next section.

The solution to Eq. (13) provides the acoustic pressure field $p(\boldsymbol{x}, t) \in \Omega$. The impulse response $h(t)$ is then calculated as,

$$h(t) = \mathcal{F}^{-1}\left[\hat{h}(f)\right] = \mathcal{F}^{-1}\left[\frac{\hat{p}(\boldsymbol{x}_h, f)}{\hat{G}(f)}\right]. \tag{15}$$

3. Numerical simulations

3.1. Numerical strategy

Before detailing the results from the numerical simulations, let us briefly summarize the numerical strategy we followed to solve the various equations in the preceding sections, in particular the ODE system in Eq. (1) to obtain the glottal volume velocity of Eq. (6), and the acoustic wave problem in Eq. (13) to obtain the impulse response of Eq. (15). Remember that with the volume velocity and the impulse response we can compute vowel sounds for various phonation regimes through the convolution in Eq. (11). All the equations in this work were solved using in-house developed numerical codes.

The ODE system in Eq. (1) was solved using the explicit fourth-order Runge–Kutta method (RK4) with a sampling frequency $f_s = 10^5$ Hz and an initial condition $(\boldsymbol{v}^{\text{T}}(0), \boldsymbol{x}^{\text{T}}(0)) = (0, 0, 0.1, 0.1)$ (see e.g., [5,24–26]). The wave problem in Eq. (6) was solved with the method of lines,

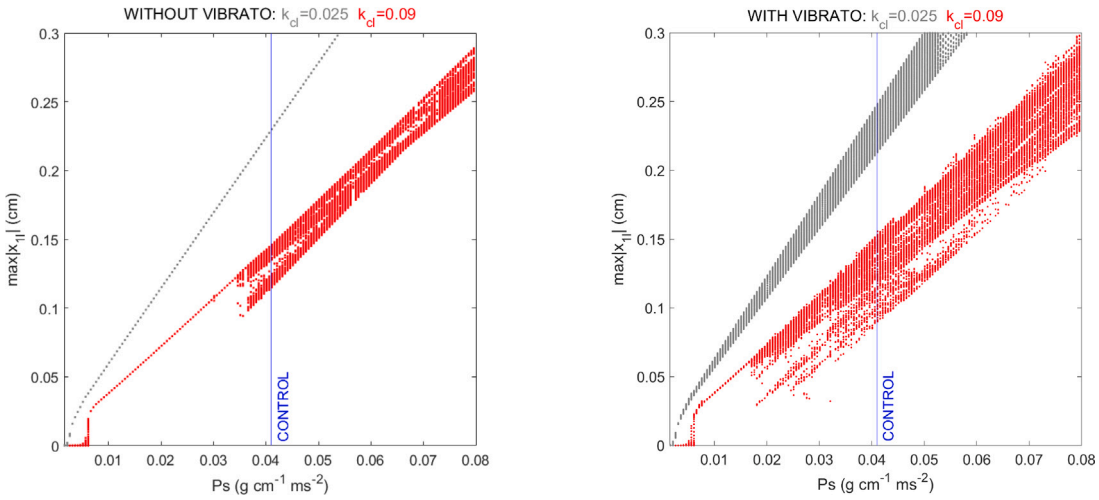

Fig. 3. Bifurcation plot without vibrato (left) and with vibrato (right) for normal ($k_{cl} = 0.025$, gray color) and abnormal ($k_{cl} = 0.09$, red color) phonation. The blue line indicates the subglottal pressure value for which chaos control simulations have been performed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

i.e., using FEM for the spatial discretization and then finite differences for the time discretization (see e.g., [59–61]). The computational domain, Ω, where we set the problem comprises the vocal tract, Ω_{VT}, plus the exterior domain, Ω_{ext}, which consists of a square cuboid of dimensions $0.3 \times 0.3 \times 0.2$ m^3. The cuboid is surrounded with a PML that has a relative reflection coefficient of 10^{-4} and a width of 0.1 m (see [61]). The rigid baffle Γ_{B} where the mouth is inserted is one of the faces of the cuboid and has dimensions 0.3×0.3 m^2. The computational domain was meshed with linear tetrahedral elements with characteristic size, h, ranging from $h \approx 0.001$ m inside the vocal tract to $h \in [0.0025, 0.005]$ m in the exterior domain and $h \approx 0.0075$ within the PML. This resulted in 3.5 to 4×10^6 elements depending on which vowel vocal tract (/ɑ/, /i/ or /u/) was considered. The weak form of Eq. (13) supplemented with the auxiliary functions and equations for the PML (see [58]) was discretized with the Galerkin FEM using piecewise linear shape functions to approximate the unknown acoustic pressure, the auxiliary variables and the test functions (see [60]). As for the time discretization of Eq. (6), a second-order finite difference central scheme was implemented for the time derivative of the acoustic pressure, while a first order central scheme was used for the time derivatives of the auxiliary variables in the PML region to introduce additional numerical dissipation that improves the performance of the PML. The sampling frequency was $f_s = 8000$ kHz to fulfill the Courant–Friedrich–Levy condition required by explicit numerical schemes.

In terms of computational cost, solving the ODE of Eq. (1) in a standard serial computing system with processor Intel® Core™ i7 3.4 GHz took only a few seconds. However, obtaining the impulse response $h(t)$ for each vowel took between 57 to 65 h with the same computer.

3.2. Regular, chaotic and controlled phonation

For the simulations in the following sections, we have used the physical and geometrical parameters from Table 1. Hereafter, the dimensions of the various variables will not be written explicitly (except in the figures), but will correspond to appropriate combinations of length, $[L] = \text{cm}$, time, $[T] = \text{ms}$ and mass, $[M] = \text{g}$, dimensions.

Let us begin by presenting bifurcation plots in Fig. 3 for the maximum amplitude of the lower mass, $\max|x_{1l}|$, as a function of the subglottal pressure, P_s, for the cases without vibrato (left subfigure) and with vibrato (right subfigure). The gray curves in the subfigures correspond to the healthy phonation case, when $k_{cl} = 0.025$, see Fig. 1a, while the red ones are those obtained when the coupling stiffness is increased to $k_{cl} = 0.09$, see Fig. 1b. For the healthy case without vibrato, the onset of vocal fold self-oscillations already takes place for very small values of P_s. A periodic motion of growing amplitude is established with increasing subglottal pressure, as indicated by the straight gray line in Fig. 3 (left). As can be seen from the red curve in the same figure, the situation is very different if the coupling stiffness is too large. First, it is obviously more difficult to set the vocal folds in motion. For small values of the subglottal pressure they stay still until $P_s \sim 0.005$ is reached. A periodic motion is then established that has a smaller amplitude than that of healthy phonation, which is logical given the higher value of the coupling stiffness. However, when $P_s \sim 0.036$, the system transitions from periodic to chaotic motion and remains there except for small windows of quasiperiodic motion. The vertical blue line in the figure labeled *control* indicates the value of $P_s \sim 0.041$ to which the control strategy will be applied in subsequent sections.

In the case with vibrato (see Fig. 3, right), the gray line of healthy phonation exhibits a non-constant amplitude for a given value of P_s, as a consequence of the modulating term of Eq. (4). The range of amplitude variation widens with P_s, its central value being given by the gray line in Fig. 3 (left). For $k_{cl} = 0.09$ (red line in Fig. 3, right), the onset of vocal fold self-oscillations also takes place for $P_s \sim 0.005$. Next, the motion becomes periodic with hardly any noticeable amplitude modulation and suddenly transitions to a chaotic regime for $P_s \sim 0.018$, much earlier than in the case without vibrato. As for the latter, the vertical blue line in the figure specifies the value at which the chaos control strategy will be tested.

In the following, let us fix the subglottal pressure at $P_s = 0.041$. Further information on the system dynamics can be gathered by looking at the phase space plots for the motion of the two masses, as well as for the glottal volume velocity. We first address the case without vibrato in Fig. 4. The left and central subfigures respectively represent the trajectories in phase space, (x_{il}, v_{il}), $i = 1, 2$, for masses m_{1l} and m_{2l}. As observed, for healthy phonation (gray lines in the subfigures) both masses exhibit apparent period-1 orbits. However, if the coupling stiffness is set to $k_{cl} = 0.09$ the trajectories of the masses become trapped in a low dimensional strange attractor (red lines). If the pacemaker is activated, the motion becomes regular again (blue lines). Obviously, one cannot recover the amplitude of the oscillation of the regular motion because the stiffness of the injured coupling spring is greater than that of the healthy one, but, most importantly, periodic or quasiperiodic self-oscillations of the vocal-folds can be restored. In the right subfigure of Fig. 4 we present results for the glottal volume velocity, U_a in Eq. (6), in time delay coordinates, i.e., $(U_a(t), U_a(t+\tau))$ with τ being the time lag. As previously explained, U_a is our variable of interest for voice generation as it plays the role of the acoustic source

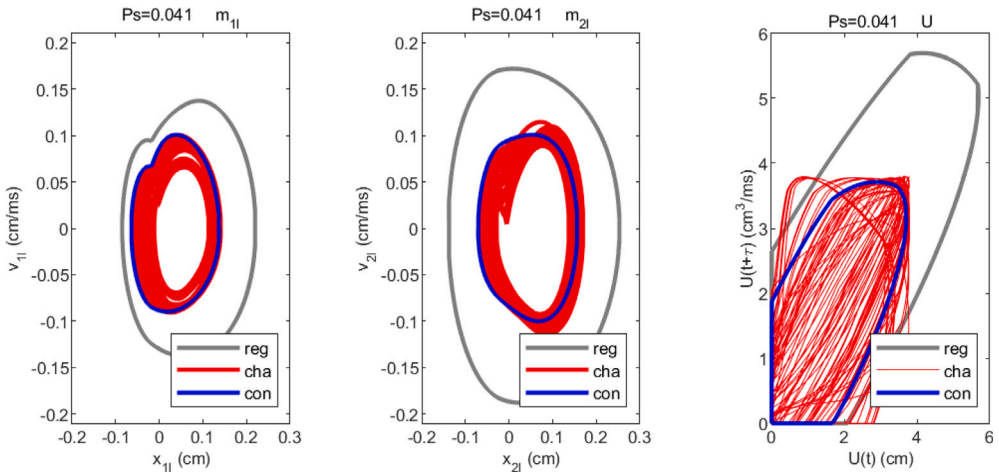

Fig. 4. Phase space plots for vocal folds without vibrato for regular phonation (gray lines), chaotic phonation (red lines) and controlled phonation (blue lines). Left: phase space of v_{1l} versus x_{1l} for m_{1l}. Center: phase space of v_{2l} versus x_{2l} for m_{2l}. Right: phase space for the glottal volume velocity in time delay coordinates $U_a(t+\tau)$ versus $U_a(t)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

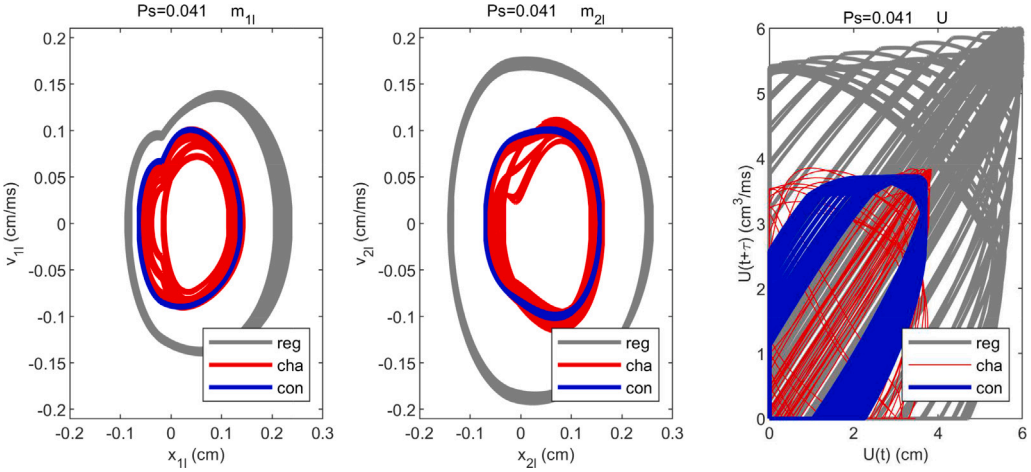

Fig. 5. Phase space plots for vocal folds with vibrato for regular phonation (gray lines), chaotic phonation (red lines) and controlled phonation (blue lines). Left: phase space of v_{1l} versus x_{1l} for m_{1l}. Center: phase space of v_{2l} versus x_{2l} for m_{2l}. Right: phase space for the glottal volume velocity in time delay coordinates $U_a(t+\tau)$ versus $U_a(t)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

term at the glottis. It is seen in the figure that its behavior is similar to that found for the motion of the two masses in phase space. In the case of healthy phonation, a closed orbit (gray line) is obtained. Note that U_a is zero at various points corresponding to the time intervals where the vocal folds collide and remain closed during a self-oscillation cycle. When the motion becomes chaotic, so does the glottal volume flux velocity, a problem that is solved when the pacemaker is activated (blue line). Though regular again, the amplitude of U_{con} is smaller than that of healthy regular phonation, U_{reg}, which will result in a decrease of the acoustic pressure intensity of the generated sound. Nevertheless, what is critical for voice intelligibility is that the vocal folds self-oscillations recover quasi-periodicity (see sections below), which is achieved by the pacemaker. In Fig. 5 we present analogous results to those in Fig. 4 but considering vibrato. The results are similar, except for the fact that now the vocal fold oscillation becomes modulated by the stiffness time dependent term of Eq. (4). As a consequence, the trajectories in the phase space of the two masses have a periodicity greater than one, even in the case of healthy phonation. This very evident in the subfigures, as well as the appearance of strange attractors for the chaotic motion and the efficiency of the pacemaker to bring motion back to normal.

We can gain some more insight into the behavior of U_a in the three analyzed regimes by calculating its power spectrum $S_{U_aU_a}$. This has been plotted in Fig. 6 left for the case without vibrato. A log–log scale has been used to help illustrate some of the typical characteristics of a regular glottal volume velocity and compare it with chaotic and controlled ones. The same color coding is used as in the previous figures. The frequency axis has been set to range from 100 Hz to 10 kHz since we are mainly interested in the spectrum beyond the fundamental frequency, which in this case is $f_0 \approx 149$ Hz (first gray peak). The typical linear decay of $S_{U_aU_a}$ can be seen in the figure. It can be also observed that the power spectrum for regular phonation essentially consists of the fundamental frequency and its harmonics, together with a pair of subharmonics for each harmonic (those of f_0 are below 100 Hz and not shown in the figure). The contributions of the subharmonics is negligible, as their peaks are ~70 dB smaller than those of the harmonics. The situation changes drastically when the vocal fold motion becomes chaotic. The peaks of the fundamental frequency and first harmonics are still present in the spectrum, but the latter flattens out at higher frequencies. The contributions of the frequencies between harmonics are now very important and make a significant impact on the energy of the glottal volume velocity. When the control is activated, the essence of the regular motion spectrum is restored (compare the gray and blue lines). The fundamental frequency and its harmonics are recovered very accurately, with a negligible contribution from the intermediate frequencies. This supports the observations made for Fig. 4 right and shows the effectiveness of the control mechanism.

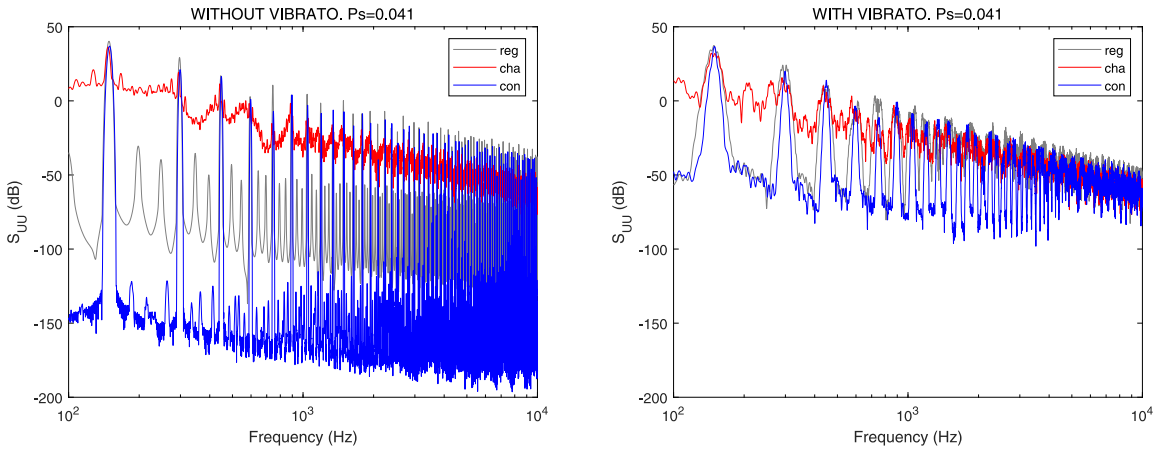

Fig. 6. Power spectrum $S_{U_aU_a}$ of the glottal volume flow without vibrato (left) and with vibrato (right) for regular phonation (gray lines), chaotic phonation (red lines) and controlled phonation (blue lines). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 6 right presents the power spectrum of the glottal volume velocity when vibrato is activated. In this case we no longer have a fundamental frequency, but, as mentioned above, several close to it, resulting in broader peaks in the spectrum even for regular phonation. As would also be expected, the level of intermediate frequencies is much higher than that observed in the absence of vibrato. The differences between the regular and chaotic cases (gray vs. blue lines) are again very visible, with important contributions from the frequencies in between peaks. For frequencies above 1.5 kHz a rather broadband spectrum is observed, even for healthy phonation. The comparison between the blue and gray lines again shows the effectiveness of the chaos control strategy.

3.3. Production of vowel sounds

3.3.1. Vocal tract transfer functions and impulse responses of vowels /ɑ/, /i/ and /u/

The VTTFs, $\hat{h}(f)$, and VT impulse responses, $h(t)$, of vowels /ɑ/, /i/ and /u/ have been calculated following the numerical strategy in section 3.1 to see how chaotic phonation affects their sound and how the pacemaker can modify it (see next subsection). The VT geometries of each vowel were taken from [38], which adapts the MRI-based VTs in [37] by removing the subglottal tube and part of the face and by finally setting the resulting VT on a flat flange that emulates the human head. The VTTFs for each vowel, $\hat{h}(f)$, have been plotted in Fig. 7. It is to be noticed that vowels are produced by the excitation of resonances of the air volume inside the vocal tract. These resonances reflect as peaks in $\hat{h}(f)$ and are usually named formants in the voice and speech research communities. The first two formants in $\hat{h}(f)$, the so-called F1 and F2, allow listeners to distinguish one vowel sound from another. Note in the figure that they have very different values for the three vowels /ɑ/, /i/ and /u/ under examination. Formants below 5–6 kHz basically correspond to planar wave propagation modes, while those at higher frequencies exhibit a three-dimensional behavior. The strong antiresonance at 5–6 kHz is caused by local resonances of the lateral cavities of the VT (piriform sinuses and valleculae), see Fig. 2 and [62,63].

3.3.2. Generation of normal and abnormal vowel sounds: efficiency of the chaos control strategy

Once we have the glottal volume velocity in the time and frequency domains, namely $U_a(t)$ and $\hat{U}_a(f)$ for $a = \mathrm{reg}, \mathrm{cha}, \mathrm{con}$, and the VTTFs and VT impulse responses, $\hat{h}(f)$ and $h(t)$, for the three vowels /ɑ/, /i/ and /u/, we can compute their acoustic pressure power spectrum $S_{pp}(f)$ and acoustic pressure time evolution from the results in Section 2.3.

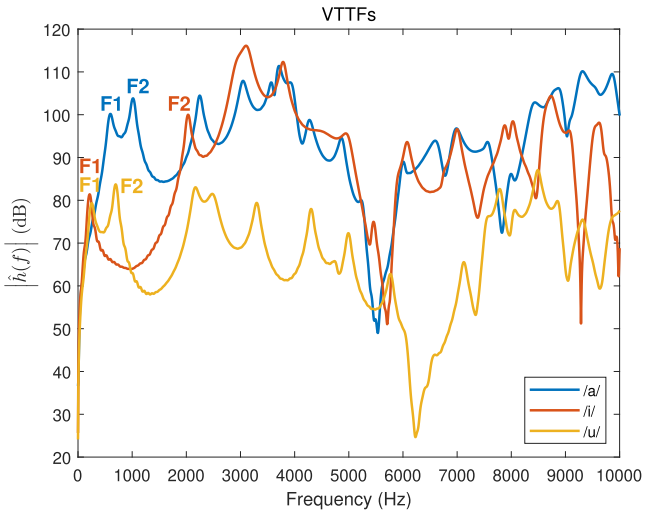

Fig. 7. Vocal tract transfer functions (VTTF), $\hat{h}(f)$, for vowels /ɑ/, /i/ and /u/.

The VTTF essentially acts as a filter for the acoustic source term, $\hat{U}_a(f)$, shaping the acoustic pressure spectrum of the vowel. In Fig. 8 we have plotted the power spectra of the acoustic pressure for the three vowels in the cases without (left column) and with (right column) vibrato. The $S_{pp}(f)$ resulting from chaotic (red lines) and controlled (blue lines) phonation are presented. Regular phonation is not considered in this figure to facilitate the comparison between the chaotic and controlled cases. The VTTF of each vowel has been superimposed on each figure to see how they conform the power spectrum. By looking at Fig. 8, similar conclusions can be drawn to those given for the glottal volume velocity in Fig. 6. If we focus first on the left column (without vibrato), it can be seen that the chaos controlling strategy produces an acoustic power spectrum that is driven by the fundamental frequency and its harmonics, while that from chaotic phonation has very significant contributions from a broadband range of frequencies that will clearly affect the produced sound. It is remarkable that the controlling strategy works over the entire frequency range, up to 10 kHz. All these considerations apply to the three vowels, as shown in the subfigures of the left column. In the case of vibrato (right column) and for the controlled case, the harmonics of the fundamental frequency dominate the spectrum up to ~1.5 kHz and begin to loose influence up to ~5 kHz, where the spectrum becomes broadband. Logically, this can be also

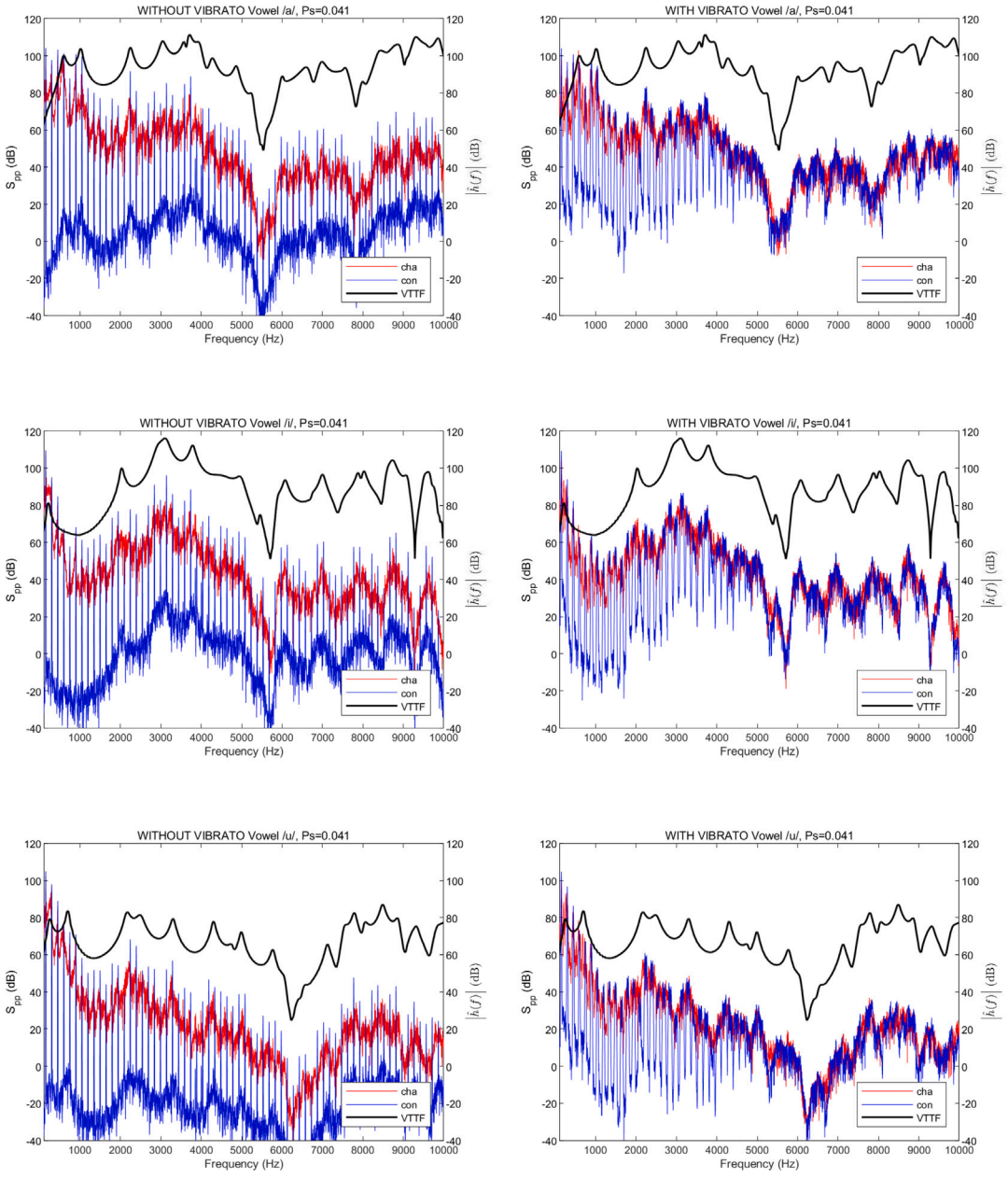

Fig. 8. Power spectrum of the acoustic pressure for vowels /ɑ/, /i/ and /u/ without vibrato (left column) and with vibrato (right column). The red lines correspond to chaotic phonation and the blue ones to controlled one. The VTTF, $\hat{h}(f)$, of each vowel has been superimposed in the figures to see how they shape the acoustic pressure spectrum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

observed in the plot for the glottal volume velocity with vibrato (Fig. 6), although the log–log scale used in that figure makes it less obvious. Achieving control up to 5 kHz in the case of vibrato is notable as this is the critical frequency range for voice sound identification, while higher frequencies are more related to voice quality [39,64,65].

To better perceive the differences in the generated vowels for regular, chaotic and controlled phonation, in Fig. 9 we present, for the case without vibrato, the time evolution and spectrogram of the acoustic pressure of each vowel for a sequence in which the vocal folds transition from a regular to a chaotic and then to a controlled motion. It can be observed in the time evolution of each vowel how the acoustic pressure due to regular phonation completely changes its pattern when chaotic phonation begins. On the one hand, the amplitude decreases markedly due to increased coupling stiffness, but also the shape of the signal is completely distorted. When the controller is activated, the shape of the acoustic pressure due to regular phonation is mostly recovered, but not the amplitude, since, as mentioned above, the coupling stiffness remains the same. In the spectrograms of the vowels, the harmonics can be clearly identified as horizontal lines, their amplitude being shaped by the VTTFs in Fig. 7. In fact, a vertical cut in the spectrogram provides the power spectra of Fig. 8 left. Three videos with audio are provided in the supplementary material for vowels /ɑ/, /i/ and /u/, which correspond to the top, middle and bottom subfigures of Fig. 9. Listening to these audios shows how the vowels sound very clear when phonation is regular, become almost unrecognizable for chaotic phonation, and can be recognized again

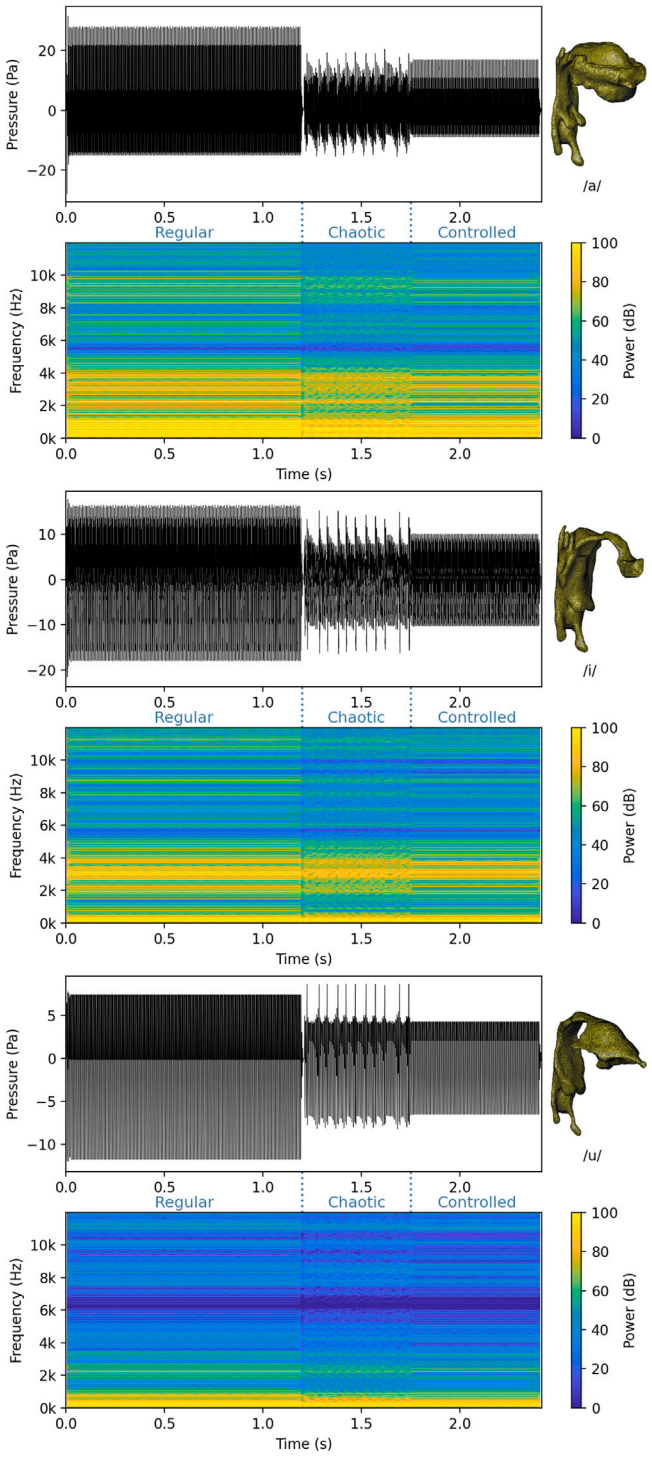

Fig. 9. Time evolution and spectrogram for vowels /ɑ/, /i/ and /u/ in the regular, chaotic and controlled phonation regimes without vibrato. Access the supplementary material to listen to the corresponding audio files.

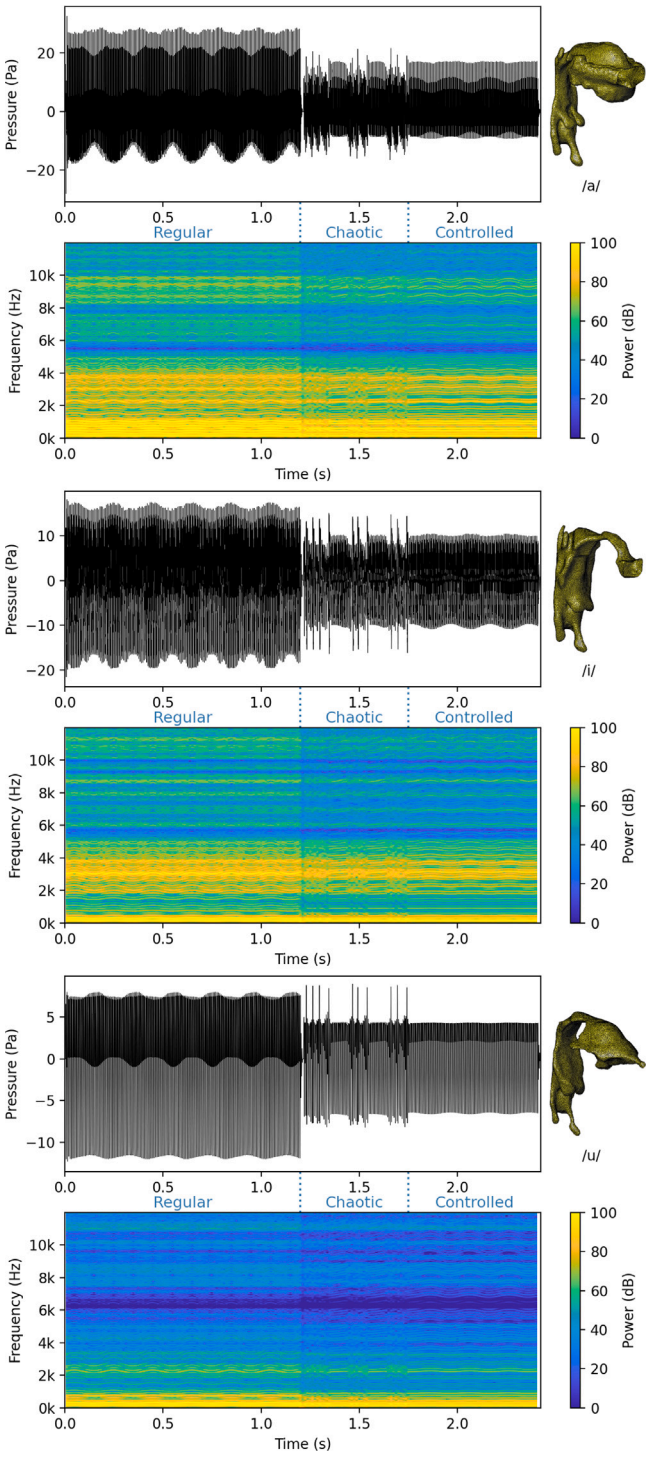

Fig. 10. Time evolution and spectrogram for vowels /ɑ/, /i/ and /u/ in the regular, chaotic and controlled phonation regimes wit vibrato. Access the supplementary material to listen to the corresponding audio files.

when the controller is activated. The respective names of the files are Vowel_m_reg_cha_without_vibrato.mp4, with m = a,i,u.

In Fig. 10 we present the analogous results of Fig. 9 when vibrato is considered. The effect of the latter is clearly seen in the time evolution of the acoustic pressure as a smooth modulation of the amplitude and also as small ripples in the harmonics of the spectrogram. The corresponding audio files in the supplementary material are now termed Vowel_m_reg_cha_with_vibrato.mp4, with m = a,i,u, and listening to them shows how the vibrato results in a less robotic vowel sound with a more pleasant sound cadence.

4. Conclusions

In this work we have investigated how an ideal pacemaker could improve vowel generation quality for unhealthy phonation due to excessive VF stiffness. To this end, a symmetric two-mass model of

the VFs, including the pacemaker, has been derived which considers several types of nonlinearities, namely the cubic restoring forces of the VFs, the collisions of the VFs, and the flow excitation. In addition, the model offers the possibility to include a vibrato effect by allowing the stiffness of the lower masses to vary with time. The glottal volume velocity for the cases of regular (healthy), chaotic (unhealthy) and controlled phonation has been computed and used in combination with the impulse response of the VTs of vowels /ɑ/, /i/ and /u/ to generate the corresponding sounds. Vowel impulse responses have been obtained by solving the wave equation with FEM for MRI-based vocal tract geometries excited with a Gaussian pulse.

As a result of the simulations we conclude the following. In the case without vibrato, the acoustic pressure spectrum of the generated vowels is governed by the fundamental frequency and its harmonics for regular phonation. However, for chaotic phonation the spectrum is broadened and the vowels become unrecognizable, as clearly perceived in the audio files provided as supplementary material. When the pacemaker is activated it is not possible to recover the amplitude of healthy phonation because the stiffness of the VFs remains higher than usual, but the quasi-periodicity of the VF motion is restored and the vowels can again be clearly identified. Similar conclusions apply to the case of vibrato, although there is no fundamental frequency that drives it, since it oscillates with time. The acoustic spectrum of vowels only shows a harmonic-type behavior for lower to mid frequencies. However, the pacemaker is also very efficient in this case and one can listen in the supplementary material how the vowels become recognizable when it is activated. The modulating effect of vibrato is also very clearly perceived.

Ongoing work by the authors is now focused on extending the chaos control strategy to asymmetric mass models, as well as including other effects in symmetric mass models to cover a wide range of voice disorders. The present work provides a direct way to test the performance of phonation pacemakers for these new scenarios by listening to numerically produced vowels for different regimes of glottal volume velocities.

CRediT authorship contribution statement

Oriol Guasch: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Marc Freixes:** Writing – review & editing, Validation, Software, Formal analysis. **Marc Arnela:** Writing – review & editing, Validation, Software, Formal analysis. **Annemie Van Hirtum:** Writing – review & editing, Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The first three authors thank the support of the project FEMVoQ (PID2020 120441GB-I00 AEI/10.13039/501100011033) from the Spanish Ministerio de Ciencia e Innovación and that of the Catalan Government (Departament de Recerca i Universitats) for the grant 2021 SGR 01396 given to the HER group. The fourth author would like to acknowledge the Full3DTalkingHead project (ANR-20-CE23-0008-03) from l'Agence Nationale de la Recherche, France.

References

[1] Van den Berg J. Myoelastic-aerodynamic theory of voice production. J Speech Hear Res 1958;1:227–44.
[2] Titze I. The myoelastic aerodynamic theory of phonation, national centre for voice and speech, iowa city. 2006.
[3] Ishizaka K, Flanagan J. Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell Syst Tech J 1972;51:1233–68.
[4] Lucero JC. Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region. J Acoust Soc Am 1993;94:3104–11.
[5] Story BH, Titze IR. Voice simulation with a body-cover model of the vocal folds. J Acoust Soc Am 1995;97:1249–60.
[6] Cisonni J, Van Hirtum A, Pelorson X, Lucero J. The influence of geometrical and mechanical input parameters on theoretical models of phonation. Acta Acust United Ac 2011;97:291–302.
[7] Erath BD, Zanartu M, Stewart KC, Plesniak MW, Sommer DE, Peterson SD. A review of lumped-element models of voiced speech. Speech Commun 2013;55:667–90.
[8] de Oliveira Rosa M, Pereira JC, Grellet M, Alwan A. A contribution to simulating a three-dimensional larynx model using the finite element method. J Acoust Soc Am 2003;114:2893–905.
[9] Vampola T, Horáček J, Klepáček I. Computer simulation of mucosal waves on vibrating human vocal folds. Biocybern Biomed Eng 2016;36:451–65.
[10] Jiang W, Zheng X, Xue Q. Computational modeling of fluid–structure–acoustics interaction during voice production. Front Bioeng Biotechnol 2017;5:7.
[11] Schoder S, Weitz M, Maurerlehner P, Hauser A, Falk S, Kniesburges S, Döllinger M, Kaltenbacher M. Hybrid aeroacoustic approach for the efficient numerical simulation of human phonation. J Acoust Soc Am 2020;147:1179–94.
[12] Döllinger M, Zhang Z, Schoder S, Šidlof P, Tur B, Kniesburges S. Overview on state-of-the-art numerical modeling of the phonation process. Acta Acust 2023;7:25.
[13] Murray PR, Thomson SL. Synthetic, multi-layer, self-oscillating vocal fold model fabrication. J Vis Exp 2011;e3498.
[14] Murray PR, Thomson SL. Vibratory responses of synthetic, self-oscillating vocal fold models. J Acoust Soc Am 2012;132:3428–38.
[15] Van Hirtum A, Bouvet A, Tokuda I, Pelorson X. Dynamic vibration mode decomposition of auto-oscillating vocal fold replicas without and with vertical tilting. J Sound Vib 2022;516:116504.
[16] Van Hirtum A, Ahmad M, Chottin R, Pelorson X. A composite analogy to study the linear elasticity of a pressurised latex tube with application to a mechanical vocal fold replica. Int J Appl Mech 2023.
[17] Bodaghi D, Xue Q, Zheng X, Thomson S. Effect of subglottic stenosis on vocal fold vibration and voice production using fluid–structure–acoustics interaction simulation. Appl Sci 2021;11:1221.
[18] Falk S, Kniesburges S, Schoder S, Jakubaß B, Maurerlehner P, Echternach M, Kaltenbacher M, Döllinger M. 3D-FV-FE aeroacoustic larynx model for investigation of functional based voice disorders. Front Physiol 2021;12:616985.
[19] Pickup BA, Thomson SL. Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models. J Biomech 2009;42:2219–25.
[20] Bouvet A, Tokuda I, Pelorson X, Van Hirtum A. Influence of level difference due to vocal folds angular asymmetry on auto-oscillating replicas. J Acoust Soc Am 2020;147:1136–45.
[21] Ahmad M, Pelorson X, Fernández I, Guasch O, Van Hirtum A. Low-strain effective Young's modulus model and validation for multi-layer vocal fold-based silicone specimens with inclusions. J Appl Phys 2022;131.
[22] Ahmad M, Pelorson X, Guasch O, Fernández AI, Van Hirtum A. Modelling and validation of the non-linear elastic stress–strain behaviour of multi-layer silicone composites. J Mech Behav Biomed Mat 2023;139:105690.
[23] van Hirtum A, Ahmad M, Chottin R, Guasch O, Pelorson X. Experimental study of the influence of a rectangular vocal folds inclusion on their auto-oscillation. In: Forum acusticum. 2023.
[24] Steinecke I, Herzel H. Bifurcations in an asymmetric vocal-fold model. J Acoust Soc Am 1995;97:1874–84.
[25] Jiang JJ, Zhang Y, Stern J. Modeling of chaotic vibrations in symmetric vocal folds. J Acoust Soc Am 2001;110:2120–8.
[26] Zhang Y, Jiang J, Rahn DA. Studying vocal fold vibrations in parkinson's disease with a nonlinear model. Chaos 2005;15.
[27] Zhang Y, Jiang JJ. Chaotic vibrations of a vocal fold model with a unilateral polyp. J Acoust Soc Am 2004;115:1266–9.
[28] Santos J, Montalvao J, Santos I. Improved model for vocal folds with a polyp with potential application. In: INTERSPEECH. p. 1386–90.

[29] Guasch O, Van Hirtum A, Fernández AI, Arnela M. Controlling chaotic oscillations in a symmetric two-mass model of the vocal folds. Chaos Solit Fractals 2022;159:112188.
[30] Tereshko V, Chacón R, Preciado V. Controlling chaotic oscillators by altering their energy. Phys Lett A 2004;320:408–16.
[31] Tereshko V. Control and identification of chaotic systems by altering their energy. Chaos Solit Fractals 2009;40:2430–46.
[32] de Souza SL, Caldas IL, Viana RL. Damping control law for a chaotic impact oscillator. Chaos Solit Fractals 2007;32:745–50.
[33] Fraile R, Saenz-Lechon N, Godino-Llorente J, Osma-Ruiz V, Fredouille C. Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex. Folia Phoniatr Logop 2009;61:146–52.
[34] Muhammad G, Mesallam TA, Malki KH, Farahat M, Mahmood A, Alsulaiman M. Multidirectional regression (mdr)-based features for automatic voice disorder detection. J Voice 2012;26:817–e19.
[35] Titze IR, Story B, Smith M, Long R. A reflex resonance model of vocal vibrato. J Acoust Soc Am 2002;111:2272–82.
[36] Arroabarren I, Carlosena A. Voice production mechanisms of vocal vibrato in male singers. IEEE Trans Audio Speech Lang Process 2006;15:320–32.
[37] Aalto D, Aaltonen O, Happonen R-P, Jääsaari P, Kivelä A, Kuortti J, Luukinen J-M, Malinen J, Murtola T, Parkkola R, Saunavaara J, Soukka T, Vainio M. Large scale data acquisition of simultaneous MRI and speech. Appl Acoust 2014;83:64–75.
[38] Arnela M, Dabbaghchian S, Blandin R, Guasch O, Engwall O, Van Hirtum A, Pelorson X. Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds. J Acoust Soc Am 2016;140:1707–18.
[39] Freixes M, Arnela M, Socoró JC, Alías F, Guasch O. Glottal source contribution to higher order modes in the finite element synthesis of vowels. Appl Sci 2019;9:4535.
[40] Takemoto H, Mokhtari P, Kitamura T. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method. J Acoust Soc Am 2010;128(6):3724–38.
[41] Shankar R, Ghosh TK, Spontak RJ. Electroactive nanostructured polymers as tunable actuators. Acta Math 2007;19:2218–23.
[42] Lamberti A, Di Donato M, Chiappone A, Giorgis F, Canavese G. Tunable electromechanical actuation in silicone dielectric film. Smart Mater Struct 2014;23:105001.
[43] Armstrong DP, Spontak RJ. Designing dielectric elastomers over multiple length scales for 21st century soft materials technologies. Rubber Chem Technol 2017;90:207–24.
[44] Hames NT, Balsbough D, Yan J, Wu S, Zuo X, Spontak RJ. Tunable thermoplastic elastomer gels derived from controlled-distribution triblock copolymers with crystallizable endblocks. Mater Horiz 2023;10:4968–75.
[45] Boretos J, Boretos S. Biomedical elastomers. In: Handbook of biomaterial properties. 1998, p. 302–39.
[46] Balmforth NJ, Frigaard IA, Ovarlez G. Yielding to stress: recent developments in viscoplastic fluid mechanics. Annu Rev Fluid Mech 2014;46:121–46.
[47] Coussot P. Yield stress fluid flows: A review of experimental data. J Non-Newton Fluid Mech 2014;211:31–49.
[48] Prost J, Jülicher F, Joanny J-F. Active gel physics. Nat Phys 2015;11:111–7.
[49] N'Gouamba E, Goyon J, Coussot P. Elastoplastic behavior of yield stress fluids. Phys Rev Fluids 2019;4:123301.
[50] Yang B, Baines R, Shah D, Patiballa S, Thomas E, Venkadesan M, Kramer-Bottiglio R. Reprogrammable soft actuation and shape-shifting via tensile jamming. Sci Adv 2021;7:eabh2073.
[51] Romero RG, Colton MB, Thomson SL. 3D-printed synthetic vocal fold models. J Voice 2021;35:685–94.
[52] Greenwood TE, Thomson SL. Embedded 3D printing of multi-layer, self-oscillating vocal fold models. J Biomech 2021;121:110388.
[53] Arnela M, Guasch O, Alías F. Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations. J Acoust Soc Am 2013;134(4):2946–54.
[54] Guasch O, Arnela M, Codina R, Espinoza H. A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production. Acta Acust United Ac 2016;102:94–106.
[55] Arnela M, Dabbaghchian S, Guasch O, Engwall O. MRI-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs. IEEE/ACM Trans Audio Speech Lang Process 2019;27:2173–82.
[56] Dabbaghchian S, Arnela M, Engwall O, Guasch O. Simulation of vowel-vowel utterances using a 3D biomechanical-acoustic model. Int J Numer Meth Biomed Eng 2021;37:e3407.
[57] Švancara P, Horáček J. Numerical modelling of effect of tonsillectomy on production of czech vowels. Acta Acust United Ac 2006;92:681–8.
[58] Kaltenbacher B, Kaltenbacher M, Sim I. A modified and stable version of a perfectly matched layer technique for the 3-d second order wave equation in time domain with an application to aeroacoustics. J Comput Phys 2013;235:407–22.
[59] Codina R. Comparison of some finite element methods for solving the diffusion-convection-reaction equation. Comput Methods Appl Mech Engrg 1998;156:185–210.
[60] Arnela M, Guasch O. Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method. J Acoust Soc Am 2013;133(6):4197–209.
[61] Arnela M, Blandin R, Dabbaghchian S, Guasch O, Alías F, Pelorson X, Van Hirtum A, Engwall O. Influence of lips on the production of vowels based on finite element simulations and experiments. J Acoust Soc Am 2016;139:2852–9.
[62] Takemoto H, Adachi S, Mokhtari P, Kitamura T. Acoustic interaction between the right and left piriform fossae in generating spectral dips. J Acoust Soc Am 2013;134:2955–64.
[63] Vampola T, Horáček J, Švec JG. Modeling the influence of piriform sinuses and valleculae on the vocal tract resonances and antiresonances. Acta Acust United Ac 2015;101:594–602.
[64] Monson BB, Lotto AJ, Ternstrom S. Detection of high-frequency energy changes in sustained vowels produced by singers. J Acoust Soc Am 2011;129(4):2263–8.
[65] Monson BB, Hunter EJ, Lotto AJ, Story BH. The perceptual significance of high-frequency energy in the human voice. Front Psychol 2014;5:587.