forum acusticum 2026
8-12 September · Graz, Austria

NON-ITERATIVE ENERGY BALANCED SCHEME FOR VOCAL AND REED INSTRUMENT SYNTHESIS

Thomas Risse[1], Stefan Bilbao[1], Thomas Hélie[1]

[1] *STMS lab, IRCAM-CNRS-Sorbonne University, Paris, France*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

An energy-balanced linearly implicit scheme for a class of systems including nonlinear dissipation laws is proposed. It is based on previous work on virtual analog simulations and bowed string simulations, and allows an efficient update. Specifically, the scheme is geared towards real-time synthesis of reed instruments and voice.
In order to include conservative nonlinearities (derived from a non-quadratic potential energy), the proposed method is built as an extension of an already existing scalar auxiliary variable scheme, handling the nonlinearity in the conservative part of the dynamics.
The scheme is framed in terms of the port-Hamiltonian framework. It is shown that the method satisfies a discrete equivalent to a continuous-time power-balance law, with positive dissipated power, thus ensuring stability.
Simplified models are used, focusing on the nonlinearities of interest. Simulation codes are available in a public repository.

Keywords: *physical modeling, real-time synthesis, voice, port-Hamiltonian systems*

1. Introduction

Physical models of musical instruments are almost invariably nonlinear. Some nonlinearities are conservative, in the sense that they derive from an energy potential, conserved in the absence of losses. These include contact potentials and geometric nonlinearities for mechanical systems, and are often distributed. On the other hand, self-oscillating instruments also rely on other types of nonlinearity, generally dissipative in nature, lumped and zero-dimensional (e.g. a friction law of a bow string interaction, or dissipation of the kinetic energy of the jet formed in a narrow channel with rapidly varying geometry). General descriptions of nonlinearities in musical instruments can be found in e.g. [1], [2], [3].
Real-time synthesis based on physical modelling requires the use of numerical methods that are both computationally efficient and robust (numerically stable). For conservative nonlinearities associated with a non-quadratic energy, a class of explicit Scalar Auxiliary Variable (SAV, see [4]) schemes has been proposed, and applied to models of musical instruments including geometric nonlinearities (e.g. [5]) or contact nonlinearities (e.g. [6]). These schemes have proven to be efficient enough to run simulations for large systems in real-time, such as gongs [7]. For dissipative nonlinearities in musical instruments, numerical schemes are often either (linearly-)implicit (e.g. [8], [9], [10]), or lacking in a discrete-time energy balance (e.g. [11], [12], [13]). In the context of virtual-analog simulation, [14] introduced a class of linearly-implicit schemes for dissipative systems, which preserves energetic identities in some cases. Such a scheme has been used [15] for the simulation of bowed string with linear string models.
In this work, we develop a linearly-implicit scheme with efficient update handling both conservative and dissipative nonlinearities in a power-conserving way. It is based on a SAV scheme for the conservative part of the system and on the first order method from [14] for the dissipative part. Results are presented on simplified models of clarinet and voice production.

2. Continuous time modeling

Models are presented here in the port-Hamiltonian framework, in anticipation of the interconnection of different components (e.g. exciter and resonator) through their ports, in a power-preserving way. First, a general class of systems is presented. Second, simplified mouthpiece and larynx models are recast within the proposed system structure.

2.1. Port-Hamiltonian system subclass

Port-Hamiltonian models are structured state-space representations of multi-physical systems. The struc-

Corresponding author: *thomas.risse@ircam.fr.*

ture allows for the systematic enforcing of energetic identities and enables easy power-preserving coupling of sub-systems.

Here, a sub-class of port-Hamiltonian systems corresponding to wave equations with non-quadratic potential energy and nonlinear dissipation laws is considered. The dynamics are written in terms of a state $\boldsymbol{\alpha} = [\mathbf{p}, \mathbf{q}]^\intercal$, including vectors of generalized coordinates $\mathbf{q}$ and momenta $\mathbf{p}$. The Hamiltonian (total energy) writes as a function of the state

$$H(\boldsymbol{\alpha}) = \frac{1}{2}\mathbf{p}^\intercal \mathbf{M}^{-1}\mathbf{p} + \underbrace{\frac{1}{2}\mathbf{q}^\intercal \mathbf{K}\mathbf{q} + E_{\mathrm{nl}}(\mathbf{q})}_{V(\mathbf{q})}, \tag{1}$$

with diagonal mass matrix $\mathbf{M}$, and symmetric stiffness matrix $\mathbf{K}$, both assumed positive definite. The potential energy $V(\mathbf{q})$ is the sum of a quadratic term and of a general positive function of the generalized coordinates $E_{\mathrm{nl}}(\mathbf{q})$, with $E_{\mathrm{nl}}(\mathbf{0}) = 0$.

The system dynamics, including dissipation, evolve according to

$$\underbrace{\begin{bmatrix} \dot{\mathbf{q}} \\ \dot{\mathbf{p}} \\ \mathbf{y} \end{bmatrix}}_{\mathbf{f}} = \begin{bmatrix} \mathbf{0} & \mathbf{J}_0 & \mathbf{0} \\ -\mathbf{J}_0^\intercal & -\mathbf{R}_0 & \mathbf{G} \\ \mathbf{0} & -\mathbf{G}^\intercal & -\mathbf{R}_{\mathrm{uu}} \end{bmatrix} \underbrace{\begin{bmatrix} \nabla V \\ \mathbf{M}^{-1}\mathbf{p} \\ \mathbf{u} \end{bmatrix}}_{\mathbf{e}}, \tag{2}$$

with $\mathbf{J}_0$ an arbitrary constant matrix (simply the identity matrix for canonical coordinates), $\mathbf{u}$ an input vector, and $\mathbf{y}$ a power conjugated output vector. $\mathbf{R}_0$ and $\mathbf{R}_{\mathrm{uu}}$ are positive semi-definite dissipation matrices. $\mathbf{G}$ is the input mapping matrix. While $\mathbf{R}_0$ is constant, $\mathbf{R}_{\mathrm{uu}}$ and $\mathbf{G}$ can be state and input dependent (with this dependency suppressed in (2), and in matrix forms of the models presented in this section for clarity).

By construction of the flow $\mathbf{f}$ and power-conjugated effort $\mathbf{e}$ vectors, one gets the power balance, using the chain rule and (2):

$$\tfrac{d}{dt}H + \mathbf{u}^\intercal \mathbf{y} \overset{(\mathbf{e}^\intercal \mathbf{f})}{=} -\left[(\mathbf{M}^{-1}\mathbf{p})^\intercal \;\; \mathbf{u}^\intercal\right] \begin{bmatrix} \mathbf{R}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathrm{uu}} \end{bmatrix} \begin{bmatrix} \mathbf{M}^{-1}\mathbf{p} \\ \mathbf{u} \end{bmatrix} \tag{3}$$

In the absence of an input ($\mathbf{u} = 0$), the energy is preserved (for $\mathbf{R}_0 = \mathbf{R}_{\mathrm{uu}} = 0$) or decreases owing to dissipation.

2.2. Single reed instrument

As a simple mouthpiece model for single reed instruments, consider the model in section 5 of [16]. It is a classical representation of the reed dynamics using a one degree of freedom oscillator with a collision force. The reed displacement with respect to its rest position is denoted q_{r}, the position of the lay as h_0 and the mouthpiece width as L_0, such that the area for the flow may be written as $A_{\mathrm{m}}(q_{\mathrm{r}}) = L_0[h_0 - q_{\mathrm{r}}]_+$, where $[\bullet]_+$ denotes the "positive part of." The flow in the mouthpiece is assumed steady, and incompressible. The kinetic energy below the reed is assumed fully dissipated as the flow exits the mouthpiece. Under these assumptions, and using Bernoulli's law, the volume flow Q_{m} resulting from the pressure difference $\Delta P = P_- - P_+$ (with P_- the mouth pressure and P_+ the pressure in the mouthpiece) across the reed may be written as

$$Q_{\mathrm{m}} = A_{\mathrm{m}}(q_{\mathrm{r}})\mathrm{sign}(\Delta P)\sqrt{2|\Delta P|/\rho_0}. \tag{4a}$$

In addition to the flow equation, the mouthpiece flow component is composed of an equation relating the pressure difference to the force F_{r} applied to the reed of effective area S_r by the fluid. The flow Q_{r} induced by the reed velocity v_{r} is also included, resulting in equations

$$F_{\mathrm{r}} = S_{\mathrm{r}}\Delta P, \quad Q_{\mathrm{r}} = S_{\mathrm{r}}v_{\mathrm{r}}. \tag{4b}$$

The flow component corresponding to (4) writes in matrix form as

$$\begin{bmatrix} y_0 = Q_- \\ y_1 = Q_+ \\ y_2 = F_{\mathrm{r}} \end{bmatrix} = \begin{bmatrix} -R_{\mathrm{m}} & R_{\mathrm{m}} & -S_{\mathrm{r}} \\ R_{\mathrm{m}} & -R_{\mathrm{m}} & S_{\mathrm{r}} \\ S_{\mathrm{r}} & -S_{\mathrm{r}} & 0 \end{bmatrix} \begin{bmatrix} u_0 = P_- \\ u_1 = P_+ \\ u_2 = v_{\mathrm{r}} \end{bmatrix}, \tag{5}$$

with

$$R_{\mathrm{m}}(\Delta P, A_{\mathrm{m}}) = \frac{Q_r(\Delta P, A_{\mathrm{m}})}{\Delta P}, \tag{6}$$

and where Q_- denotes the volume flow from the mouthpiece to the mouth and Q_+ the flow from the mouthpiece to the bore entrance (note that $Q_- = -Q_+$)[1]. Connection of (5) to the mechanical reed model (with reed mass m_{r} and potential $V(q_{\mathrm{r}}) = 0.5k_{\mathrm{r}}q_{\mathrm{r}}^2 + V_{\mathrm{c}}(q_{\mathrm{r}})$ associated to reed stiffness k_{r} and to a contact potential V_{c} of the form of a power law) yields the complete mouthpiece component in the form of (2)

$$\begin{bmatrix} \dot{q}_{\mathrm{r}} \\ \dot{p}_{\mathrm{r}} \\ y_0 = Q_- \\ y_1 = Q_+ \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & -R_0 & S_{\mathrm{r}} & -S_{\mathrm{r}} \\ 0 & -S_{\mathrm{r}} & -R_{\mathrm{m}} & R_{\mathrm{m}} \\ 0 & S_{\mathrm{r}} & R_{\mathrm{m}} & -R_{\mathrm{m}} \end{bmatrix} \begin{bmatrix} \nabla V \\ \frac{p_{\mathrm{r}}}{m_r} \\ u_0 = P_- \\ u_1 = P_+ \end{bmatrix}, \tag{7}$$

which then needs to be connected to an acoustic resonator (not detailed here) to build a full instrument.

2.3. Vocal

As a low order larynx model for vocal synthesis, consider the body-cover model of the vocal folds from [11]. Similar to many classical models of the larynx[2], it relies on a quasi-steady assumption for the flow, coupled to a mechanical description of the vocal folds. In that way, it resembles the mouthpiece model presented above. A difference, however, is that the vocal folds are generally represented as an assembly of mass-spring-damper systems, with several masses (3 for the body-cover model) interacting with the flow, rather than as a single degree of freedom system.

The vocal fold state includes the vector of displace-

[1] Equation (5) has A_{m} as a free parameter. After connection to the reed model in (7), A_{m} can be computed as a function of the reed displacement q_r.

[2] But contrary to previous port-Hamiltonian models of voice production (see e.g. [17], [18]) that considered unsteady glottal flow.

ments of the masses (respectively lower mass, upper mass and body mass) from their rest positions $\mathbf{q}_\mathrm{f} = [q_\mathrm{l}, q_\mathrm{u}, q_\mathrm{b}]^\intercal$, and associated momenta $\mathbf{p}_\mathrm{f}$ (where 'f' stands for fold).

The flow description is composed of two equations. The first, similar to (4a), relates the pressure drop $\Delta P = P_- - P_+$ (where this time, P_- is the sub-glottal pressure and P_+ the supra-glottal pressure) to a resulting mean flow Q_g in the glottis

$$Q_\mathrm{g} = A_\mathrm{min}(\mathbf{q}_\mathrm{f})\mathrm{sign}(\Delta P)\sqrt{2|\Delta P|/\rho_0}, \tag{8a}$$

where $A_\mathrm{min}(\mathbf{q}_\mathrm{f})$ is the minimal open area for the flow, above masses m_l and m_u. The second equation, similar to (4b), describes the forces $\mathbf{F}_\mathrm{f} = [F_\mathrm{l}, F_\mathrm{u}, F_\mathrm{b} = 0]^\intercal$ applied by the fluid pressure to the vocal folds. These forces can be written as linear combinations of the sub-glottal and supra-glottal pressures:

$$\mathbf{F}_\mathrm{f} = \mathbf{S}_{\mathrm{f}-}(\mathbf{q}_\mathrm{f})P_- + \mathbf{S}_{\mathrm{f}+}(\mathbf{q}_\mathrm{f})P_+, \tag{8b}$$

where $\mathbf{S}_{\mathrm{f}-}(\mathbf{q}_\mathrm{f})$ and $\mathbf{S}_{\mathrm{f}+}(\mathbf{q}_\mathrm{f})$ are vectors of effective surfaces depending on the current geometry of the glottis. Their expression is not detailed here but can be found by a rewriting of equations 21-23 of [11]. Note that the parametrization of (8b) is general and is suitable also for a description of forces on higher dimensional mechanical models of the vocal folds, as long as the flow is modeled as steady and incompressible.

In [11], the flow pumped by the vocal folds displacement is not included in the model, as it is mostly negligible in front of the mean flow Q_g in the glottis owing to the low range of oscillation frequencies. However, this choice results in a non-power preserving and non-mass preserving model. Hence, it is not directly compatible with the port-Hamiltonian framework. In this work, we propose to minimally modify the model to include power conjugated flows induced by the vocal fold velocity vector $\mathbf{v}_\mathrm{f} = \mathbf{M}_\mathrm{f}^{-1}\mathbf{p}_\mathrm{f}$ as

$$Q_- = -Q_g - \mathbf{S}_{\mathrm{f}-}^\intercal(\mathbf{q}_\mathrm{f})\mathbf{v}_\mathrm{f}, \quad Q_+ = Q_g - \mathbf{S}_{\mathrm{f}+}^\intercal(\mathbf{q}_\mathrm{f})\mathbf{v}_\mathrm{f}, \tag{8c}$$

with outgoing lower and upper flows Q_-, Q_+. This writing effectively restores both the mass and power balances of the model.

The flow component corresponding to (8) may be written as

$$\begin{bmatrix} y_0 = Q_- \\ y_1 = Q_+ \\ \mathbf{y}_{2,3} = \mathbf{F}_\mathrm{f} \end{bmatrix} = \begin{bmatrix} -R_\mathrm{g} & R_\mathrm{g} & -\mathbf{S}_{\mathrm{f}-}^\intercal \\ R_\mathrm{g} & -R_\mathrm{g} & -\mathbf{S}_{\mathrm{f}+}^\intercal \\ \mathbf{S}_{\mathrm{f}-} & \mathbf{S}_{\mathrm{f}+} & \mathbf{0} \end{bmatrix} \begin{bmatrix} u_0 = P_- \\ u_1 = P_+ \\ \mathbf{u}_{2,3} = \mathbf{v}_\mathrm{f} \end{bmatrix}, \tag{9}$$

with

$$R_\mathrm{g}(\Delta P, \mathbf{q}_\mathrm{f}) = \frac{Q_\mathrm{g}(\Delta P, A_\mathrm{min}(\mathbf{q}_\mathrm{f}))}{\Delta P}, \tag{10}$$

and where the difference with respect to (5) is that the velocity input and associated output force are now vectors. Also, the effective area vectors are dependent on the vocal fold geometry, whereas S_r was a constant in the case of the mouthpiece.

Similarly to the case of the mouthpiece flow component, (9) can be connected to the vocal fold mechanical model [3] through $\mathbf{u}_{2,3}$ and $\mathbf{y}_{2,3}$

$$\begin{bmatrix} \dot{\mathbf{q}}_\mathrm{f} \\ \dot{\mathbf{p}}_\mathrm{f} \\ y_0 = Q_- \\ y_1 = Q_+ \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I} & -\mathbf{R}_\mathrm{f} & \mathbf{S}_{\mathrm{f}-} & \mathbf{S}_{\mathrm{f}+} \\ \mathbf{0} & -\mathbf{S}_{\mathrm{f}-} & -R_g & R_g \\ \mathbf{0} & -\mathbf{S}_{\mathrm{f}+} & R_g & -R_g \end{bmatrix} \begin{bmatrix} \nabla V_\mathrm{f} \\ \mathbf{M}_\mathrm{f}^{-1}\mathbf{p}_\mathrm{f} \\ u_0 = P_- \\ u_1 = P_+ \end{bmatrix} \tag{11}$$

The voice model can then be completed by a vocal tract model connected to the input-output pair u_1, y_1, and controlled via the sub-glottal pressure P_- (or alternatively connected to a sub-glottal tract model connected to the remaining input-output pair).

3. Time-stepping scheme

This section presents the time-stepping scheme for (2). It is based on a SAV scheme handling the nonlinear potential forces ∇V and on an explicit evaluation of the modulated matrices $\mathbf{R}$ and $\mathbf{G}$. For brevity, the SAV transformation will not be detailed here, and the scheme is presented directly. Refer to e.g. [19] for a presentation with similar writing than used in the current paper.

3.1. Description

The state of the system is augmented by the auxiliary variable $r = \sqrt{2E_\mathrm{nl}(\mathbf{q})}$. Then, the scheme for the augmented system writes

$$\delta_{t+}\mathbf{q}^{n+\frac{1}{2}} = \mathbf{J}_0\mathbf{M}^{-1}\mathbf{p}^n, \tag{12a}$$

$$\begin{aligned} \delta_{t+}\mathbf{p}^n = & -\mathbf{J}_0^\intercal\mathbf{K}\mathbf{q}^{n+\frac{1}{2}} - \overline{\mathbf{g}}\mu_{t+}r^n \\ & -\mathbf{R}_0\mathbf{M}^{-1}\mathbf{p}^n \\ & +\overline{\mathbf{G}}\mathbf{u}^{n+\frac{1}{2}}, \end{aligned} \tag{12b}$$

$$\delta_{t+}r^n = \overline{\mathbf{g}}^\intercal\mathbf{M}^{-1}\mu_{t+}\mathbf{p}^n, \tag{12c}$$

$$\begin{aligned} \mathbf{y}^{n+\frac{1}{2}} = & -\overline{\mathbf{G}}^\intercal\mathbf{M}^{-1}\mu_{t+}\mathbf{p}^n \\ & -\overline{\mathbf{R}_\mathrm{uu}}\mathbf{u}^{n+\frac{1}{2}}, \end{aligned} \tag{12d}$$

with operators $\delta_{t+}x^n = (x^{n+1} - x^n)/dt$ and $\mu_{t+}x^n = (x^{n+1} + x^n)/2$ on time series. $\mathbf{g}(\mathbf{q}) = \mathbf{J}_0\nabla E_\mathrm{nl}(\mathbf{q})/\sqrt{2E_\mathrm{nl}(\mathbf{q})}$ is a vector resulting from the SAV transformation. Bars over $\overline{\mathbf{g}}$ and $\overline{\mathbf{R}}_\mathrm{uu}$ indicate evaluations of the state-dependent vector and matrices for the current time step. A few remarks on these evaluations:

- For the scheme to be linearly-implicit, they must be explicit functions of known value of the state and input variables,
- Independently of this choice, the discrete power-balance equation in (13) is valid.
- Ideally, all evaluations should be done at time-step $n + \frac{1}{2}$. In this case, the scheme is expected to be consistent at second order. However, this usually conflicts with the first item of this list. In this case, an evaluation can be performed at

[3] for detailed descriptions of the mechanical model dissipation matrix $\mathbf{R}_\mathrm{f}$, potential function V_f and mass matrix $\mathbf{M}_\mathrm{f}$, see section 2.2.1 of [18]

time-step n or $n-\frac{1}{2}$, but the scheme can only be consistent at first-order.

- The SAV method is known to generate artefacts related to the auxiliary variable drift. To mitigate these issues, alternative evaluations of $\overline{\mathbf{g}}$ have been proposed in previous works [6], [19], [20]. The method from [19] is directly compatible with the proposed scheme. For contact problems, constraint-based modifications of $\overline{\mathbf{g}}$ (see [6], [20]) are more suitable. It should be possible to adapt them to the proposed scheme, but this is left for future work.

3.2. Discrete power-balance

The discrete equivalent to the power-balance equation (3) for scheme (12) may be written as

$$
\begin{aligned}
&\delta_{t+}E^n + (\mathbf{u}^{n+\frac{1}{2}})^\intercal \mathbf{y}^{n+\frac{1}{2}} \\
&= -(\mathbf{M}^{-1}\mu_{t+}\mathbf{p}^n)^\intercal \mathbf{R}_0 \mathbf{M}^{-1}\mu_{t+}\mathbf{p}^n - (\mathbf{u}^{n+\frac{1}{2}})^\intercal \overline{\mathbf{R}}_{\mathrm{uu}} \mathbf{u}^{n+\frac{1}{2}}
\end{aligned} \tag{13}
$$

where the discrete energy E^n is defined in a similar manner to [19]. A CFL-like condition for stability depending on the linear part of the system only is obtained through the positivity condition of E^n. The right-hand side of the equation corresponds to the dissipated power. It is positive under the condition that $\overline{\mathbf{R}}_{\mathrm{uu}}$ is positive.

3.3. Update form

At first glance, an efficient update form for (12) is not obvious. However, one can obtain such a formulation by reworking the scheme equations slightly. One arrives at the following algorithm:

1. Compute $\mathbf{q}^{n+\frac{1}{2}}$ using (12a),
2. Compute $\overline{\mathbf{g}}$ and $\mathbf{R}_{\mathrm{uu}}$ from previous state,
3. Solve for $\mathbf{p}^{n+1}$ using the Woodbury formula,
4. Compute r^{n+1} using (12c).

Here, step 3 is the most complex. Reworking (12) yields

$$
\overline{\mathbf{A}}\mathbf{p}^{n+1} = -\mathbf{J}_0\mathbf{K}\mathbf{q}^{n+\frac{1}{2}} - \overline{\mathbf{g}}r^n + \overline{\mathbf{B}}\mathbf{p}^n + \overline{\mathbf{C}}\mathbf{u}^{n+\frac{1}{2}}, \tag{14a}
$$

with

$$
\begin{aligned}
\overline{\mathbf{A}} &= \frac{\mathbf{I}}{dt} + \left(\frac{dt}{4}\overline{\mathbf{g}}^{n+\frac{1}{2}}(\overline{\mathbf{g}}^{n+\frac{1}{2}})^\intercal\right)\mathbf{M}^{-1}, && (14b)\\
\overline{\mathbf{B}} &= \frac{\mathbf{I}}{dt} - \left(\frac{dt}{4}\overline{\mathbf{g}}^{n+\frac{1}{2}}(\overline{\mathbf{g}}^{n+\frac{1}{2}})^\intercal + \mathbf{R}_0\right)\mathbf{M}^{-1}, && (14c)\\
\overline{\mathbf{C}} &= \overline{\mathbf{G}}. && (14d)
\end{aligned}
$$

One then needs to solve a linear system in $\overline{\mathbf{A}}$, which may be written as a rank-one perturbation of a constant diagonal matrix, owing to the SAV-related term $\frac{dt}{4}\overline{\mathbf{g}}^{n+\frac{1}{2}}(\overline{\mathbf{g}}^{n+\frac{1}{2}})^\intercal$.

Both the mouthpiece and larynx models have to be connected to a resonator for simulation. Considering that an explicit scheme has been chosen for the acoustic resonator (e.g. FDTD for Webster's equation)[4], an additional relationship can be written between the pressure and volume flow at the resonator entrance as

$$
P_+^{n+\frac{1}{2}} = \overline{a} + \overline{b}Q_+^{n+\frac{1}{2}}, \tag{15}
$$

where $\overline{a}$ and $\overline{b}$ are scalars computed from the previous state of the resonator. Using this relationship together with equation (12d) for the larynx flow model, the corresponding part of the input vector $\mathbf{u}^{n+\frac{1}{2}}$ can be rewritten as

$$
P_+^{n+\frac{1}{2}} = \frac{1}{1+\overline{b}\overline{R}_{\mathrm{g}}}\left(\overline{a} + \overline{b}\left(\overline{R}_{\mathrm{g}}P_-^{n+\frac{1}{2}} + \mathbf{S}_{\mathrm{f}+}^\intercal\mathbf{M}^{-1}\mu_{t+}\mathbf{p}^n\right)\right), \tag{16}
$$

with $\overline{R}_{\mathrm{g}}$ evaluated using values of $P_+^{n-\frac{1}{2}}$ and $P_-^{n-\frac{1}{2}}$ to keep the scheme explicit (see remarks in section). Finally, step 3 of the algorithm can be modified by replacing $P_+^{n+\frac{1}{2}}$ by its value evaluated from (16), ultimately resulting in an additional first order perturbation of the matrix $\overline{\mathbf{A}}$ in (14). Using a similar construction, it is also possible to connect a sub-glottal tract to the larynx.

To summarize on the complexity of the scheme, one has to solve a linear system which is a rank-k perturbation of a constant diagonal matrix. The value of k depends on the studied model:

- If a non-quadratic potential is present, then the SAV algorithm adds a rank-1 perturbation ($k+=1$),
- If the system is connected to other systems yielding explicit input-output relationships at each time step (like in (15)), then it adds a rank(n_{c}) perturbation, where n_{c} is the number of input-output pairs connected in that way.

The mouthpiece model yields trivial inversion as it is a scalar system. The larynx model yields $k=2$ when connected to a vocal tract, and $k=3$ if it is also connected to a sub-glottal tract. Most importantly, for the larynx example, if the number of degrees of freedom of the underlying vocal fold mechanical model were to increase, the rank perturbation would stay the same, as it depends only on the number of fluid interfaces connected, which is at most 2.

4. Numerical experiments

Due to size constraint, only some results on a voice simulation are presented. Simulation codes and additional figures corresponding to the clarinet model are available in an open repository.

[4] The overall simulation fulfills a discrete power balance if the chosen scheme for the resonator itself has a discrete power balance with external power opposite with the external power of the exciter computed from (13).

A configuration similar to case A of [11] is used for the vocal folds parameters. The vocal tract, discretized using an FDTD scheme, has a straight geometry. A sub-glottal pressure ramp with maximum value 800 Pa and rise time 0.05 s is used to force the system, and the simulation is run at standard audio sample-rate fs = 44100 Hz.
Figure 1 presents the discrete power balance corresponding to equation (13) for the larynx component. The sum of the three contributions gives zeros up to the numerical precision ($\sim 10^{-16}$) at each sample, and the dissipated power is always positive.

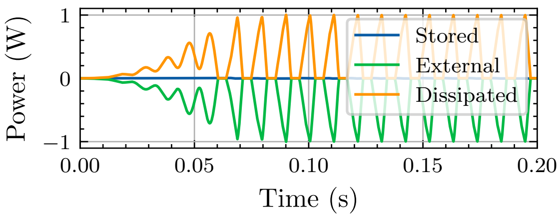

Figure 1: Discrete stored, external (negative when the system receives energy) and dissipated powers corresponding to terms of (13) for the larynx component in a self-oscillating configuration.

The vocal fold masses displacements and sup-glottal flow (input flow of the resonator) are shown in Figure 2 and correspond to the typical output of the model. Notice that the contact potential is activated during the simulation as the opening becomes negative.

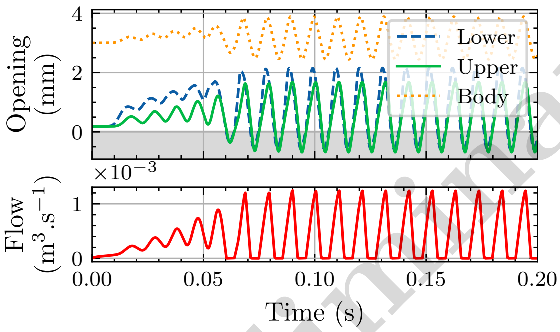

Figure 2: Vocal fold masses displacements and supra-glottal flow.

A convergence study is performed by running a set of 8 simulations for sampling frequencies $\{sr_0, 2sr_0, ..., 2^{8-1}sr_0\}$ Hz, with $sr_0 = 44100$ Hz. The simulation with the highest sampling rate is taken as the reference, and the error is computed as the L^2 error between the simulations and reference masses displacements, for the first 0.2 s of the simulation. This time window thus includes both the dissipative nonlinear effects (Bernoulli-type flow) and the potential nonlinearity (contact potential and cubic spring nonlinearities in the body-cover model). Results shown in Figure 3 assess the first order convergence of the scheme.
A second configuration is artificially created by multiplying by 10 the lower vocal fold stiffness and raising

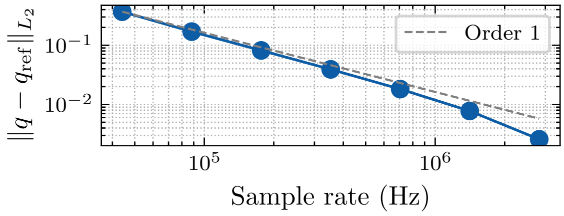

Figure 3: Convergence of the results corresponding to Figure 2 with sampling frequency.

the maximum sub-glottal pressure to 1000 Pa. This results in a period doubling oscillation regime, for which the signals are shown in Figure 4. The power balance is also satisfied up to numerical precision in this case and not shown here. First-order convergence is still observed as presented in Figure 5.

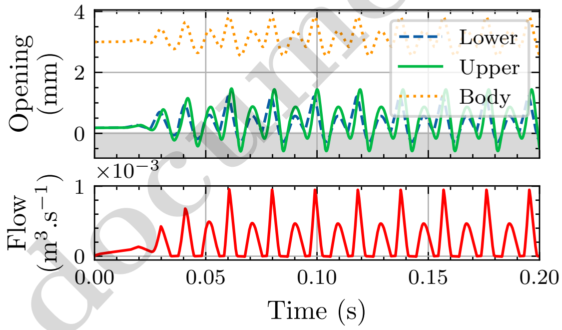

Figure 4: Vocal fold masses displacements and supra-glottal flow in a period doubling configuration.

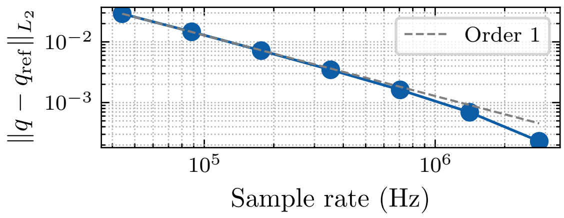

Figure 5: Convergence of the results corresponding to Figure 4 with sampling frequency.

In the two presented cases, the algorithm converges and has the right qualitative behavior at standard sampling rate $sr = 44100$ Hz. Cases were the algorithm starts to converge only at higher sampling rates were observed for configurations close to bifurcations of the system. As an example, in the second configuration studied, period-doubling is observed as the sub-glottal pressure crosses a threshold value close to 800 Pa. This threshold is observed to be higher for simulations at lower samplerate. As a result, a convergence study run with 800 Pa as the maximum sub-glottal pressure suffers as some simulations are ”stuck” in the standard periodic regime whereas the reference switches to a period-doubling regime.
Convergence was also observed for simulations of the

clarinet model. In order to properly handle reed contact with the lay, modified SAV vector evaluation needs however to be used, as presented in [6].

5. Conclusion

In this paper, we presented a scheme for a reed and vocal synthesis. It is linearly-implicit with an efficient update and satisfies a discrete equivalent to the continuous time power-balance of the studied systems. Models of single reed instrument and voice were recast in the required form for application of the scheme. The method is suitable for real-time simulations: as an example, the presented simplified voice simulation runs 200 times faster than real-time on a MacBook Pro. As a main advantage of the scheme for vocal synthesis, it should scale linearly with the size of the mechanical model of the vocal folds, under similar assumptions for the fluid flow.

6. Funding

This research was funded, in whole or in part, by l'Agence Nationale de la Recherche (ANR), project AVATARS (ANR-22-CE48-0014).

References

[1] N. H. Fletcher, "The nonlinear physics of musical instruments," *Reports on progress in physics*, vol. 62, no. 5, pp. 723–764, 1999.

[2] S. Bilbao, "The changing picture of nonlinearity in musical instruments," in *International Symp. on Musical Acoustics*, 2014.

[3] O. Thomas, C. Vergez, and C. Touzé, "Nonlinear dynamical phenomena in musical acoustics," *Handbook of Nonlinear Dynamics*, 2025.

[4] J. Shen, J. Xu, and J. Yang, "The scalar auxiliary variable (sav) approach for gradient flows," *J. Comp. Phys.*, vol. 353, pp. 407–416, 2018.

[5] R. Russo, S. Bilbao, and M. Ducceschi, "Scalar auxiliary variable techniques for nonlinear transverse string vibration," *IFAC-PapersOnLine*, vol. 58, no. 6, pp. 160–165, 2024.

[6] M. Van Walstijn, V. Chatziioannou, and A. Bhanuprakash, "Implicit and explicit schemes for energy-stable simulation of string vibrations with collisions: Refinement, analysis, and comparison," *Journal of Sound and Vibration*, vol. 569, p. 117 968, 2024.

[7] S. Bilbao, C. Webb, Z. Wang, and M. Ducceschi, "Real-time gong synthesis," in *Proc. of the 26th International Conference on Digital Audio Effects*, 2023, pp. 1–8.

[8] C. Desvages and S. Bilbao, "Physical modeling of nonlinear player-string interactions in bowed string sound synthesis using finite dierence methods," in *Proc. of the International Symp. on Musical Acoustics*, 2014.

[9] E. Gourc, C. Vergez, P.-O. Mattei, and S. Missoum, "Nonlinear dynamics of the wolf tone production," *Journal of Sound and Vibration*, vol. 516, p. 116 463, 2022.

[10] E. Matusiak and V. Chatziioannou, "A comparison of friction models for bow-string interaction based on experimental measurements," in *Proc. Stockholm Music Acoustics Conference*, 2023.

[11] B. H. Story and I. R. Titze, "Voice simulation with a body-cover model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1249–1260, 1995.

[12] M. Demoucron, "On the control of virtual violins: Physical modelling and control of bowed string instruments," Ph.D. dissertation, KTH, 2008.

[13] F. Silva, C. Vergez, P. Guillemain, J. Kergomard, and V. Debut, "Moreesc: A framework for the simulation and analysis of sound production in reed and brass instruments," *Acta Acustica united with Acustica*, vol. 100, no. 1, pp. 126–138, 2014.

[14] M. Ducceschi, S. Bilbao, and C. J. Webb, "Non-iterative schemes for the simulation of nonlinear audio circuits," in *2021 24th International Conference on Digital Audio Effects*, IEEE, 2021, pp. 25–32.

[15] R. Russo, M. Ducceschi, and S. Bilbao, "Efficient simulation of the bowed string in modal form," in *International Conference on Digital Audio Effects*, 2022, pp. 122–129.

[16] S. Bilbao, A. Torin, and V. Chatziioannou, "Numerical modeling of collisions in musical instruments," *Acta Acustica united with Acustica*, vol. 101, pp. 155–, Jan. 2015.

[17] L. A. Mora, H. Ramirez, J. I. Yuz, Y. Le Gorec, and M. Zañartu, "Energy-based fluid–structure model of the vocal folds," *IMA Journal of Mathematical Control and Information*, vol. 38, no. 2, pp. 466–492, 2021.

[18] T. Risse, T. Hélie, and F. Silva, "Voice synthesis using power-balanced simulation of a quasi-1d model of the vocal apparatus," *Frontiers in Signal Processing*, vol. 5, p. 1 525 198, 2025.

[19] T. Risse, T. Helie, and S. Bilbao, "Power-balanced drift regulation for scalar auxiliary variable methods: Application to real-time simulation of nonlinear string vibrations," in *28th International Conference on Digital Audio Effects*, IEEE, 2025, pp. 126–133.

[20] R. Russo, C. J. Webb, M. Ducceschi, and S. Bilbao, "Convergence analysis and relaxation techniques for modal scalar auxiliary variable methods applied to nonlinear transverse string vibration," in *Proc. of Meetings on Acoustics*, Acoustical Society of America, vol. 56, 2025, p. 035 007.