frontiers | Frontiers in Signal Processing

TYPE Editorial
PUBLISHED 27 October 2025
DOI 10.3389/frsip.2025.1715792

OPEN ACCESS

EDITED AND REVIEWED BY
Augusto Sarti,
Polytechnic University of Milan, Italy

*CORRESPONDENCE
Michele Ducceschi,
michele.ducceschi@unibo.it

RECEIVED 29 September 2025
ACCEPTED 07 October 2025
PUBLISHED 27 October 2025

CITATION
Hélie T, Van Walstijn M and Ducceschi M (2025) Editorial: Sound synthesis through physical modeling.
Front. Signal Process. 5:1715792.
doi: 10.3389/frsip.2025.1715792

Editorial: Sound synthesis through physical modeling

Thomas Hélie[1], Maarten Van Walstijn[2] and Michele Ducceschi[3]*

[1]STMS Laboratory, IRCAM–CNRS–Sorbonne Université, Paris, France, [2]Sonic Arts Research Centre, Queen's University Belfast, Belfast, United Kingdom, [3]NEMUS Lab, Department of Industrial Engineering, University of Bologna, Bologna, Italy

KEYWORDS

physical modeling, real-time performance, hardware-efficient computation, machine-learning–assisted modeling, multi-physics and nonlinear systems, energy-consistent numerical schemes, nonlinear systems, port-Hamiltonian systems

Editorial on the Research Topic
Sound synthesis through physical modeling

1 Introduction

Physical modeling synthesis aims to simulate sound by solving the equations governing an instrument or acoustic system's behaviour. This paradigm stands apart from signal-based methods by directly encapsulating the cause–and–effect relationships of sound production. Its appeal lies in the fidelity of transients, the mapping from parameters to sound, and the capacity to expose how instruments work. Over the past 50 years, models have grown from simple linear caricatures to treatments of strongly nonlinear, coupled systems, frictional contacts, nonlinear audio effects, nonlinear air flows, and more. That realism comes at the cost of greater numerical hurdles and, consequently, the need for schemes that remain stable and efficient. More recently, new tools have entered the picture. Machine learning is being used to tune or hybridise physical models, yielding differentiable systems that can learn while staying tied to physics. Energy-based formalisms such as port-Hamiltonian models provide a systematic way to enforce passivity in power-exchanging systems. At the same time, hardware–GPUs, FPGAs, modern CPUs–has opened up real-time simulation at scales that were out of reach only a decade ago. This Research Topic sits in that landscape. The contributions span multi-physics and nonlinear modelling, energy-consistent algorithms, machine learning assistance, and hardware acceleration, all aimed at sharpening accuracy and stability while keeping models playable in real-time. What follows is a tour of the six papers and of how they move the field forward.

2 Contributions in this issue

The articles in this Research Topic span a diverse range of musical instruments and methodological approaches, yet they are unified by an emphasis on physically grounded models and system-level rigour. Below, we summarise each article in turn and discuss its context within the broader field.

2.1 Physics-informed piano modeling

Simionato et al. present a hybrid modeling approach that fuses deep learning with traditional DSP to emulate piano sounds in a differentiable framework. Focusing on single piano notes, their method learns to synthesise the quasi-harmonic spectrum using physics-derived parametric formulas whose parameters are optimised from recorded samples. By embedding known piano acoustics (such as inharmonicity of strings and partial envelope decays) into a neural network training process, the model achieves a high degree of realism while remaining lightweight and fully interpretable. Notably, the learned synthesiser reproduces each note's partial frequencies (stretched by string stiffness) and amplitude dynamics across different key velocities. The authors report that the model generalises accurately across the piano's range, successfully capturing the inharmonicity and level of each partial. Remaining challenges include some loss of accuracy for very high-frequency partials at loud dynamics, likely due to limited training data in those regimes. Importantly, the architecture is modular and amenable to real-time use: it operates with low latency and modest computational load, making it suitable for interactive digital piano applications. This work exemplifies the trend of differentiable physical models, where neural networks are guided by physical insight–here, yielding a novel piano model that incorporates both data-driven and physics-based synthesis. It underlines how machine learning can enhance physical modeling by automating parameter tuning and system identification, all while preserving the intuitive control and explainability of classic models.

2.2 Energy-conserving woodwind simulation

Darabundit and Scavone propose a discrete port-Hamiltonian system (PHS) approach to model a single-reed woodwind instrument (such as a clarinet) in a modular, energy-consistent way. In traditional woodwind synthesis, the nonlinear reed excitation and the acoustic resonator (bore and toneholes) have often been modeled separately with methods like digital waveguides or finite-difference schemes. This article instead formulates each major component–the reed, the air column, and an open/closed tonehole–as interconnected subsystems in the port-Hamiltonian framework. By doing so, the authors ensure that energy flow between components is explicitly tracked and conserved, conferring numerical stability and physical interpretability (power balance) to the complete model. The paper presents a number of contributions: a linearly implicit integration scheme for the beating reed dynamics using an energy quadratization method to handle the nonlinear collision force (Hunt–Crossley model) coupled with nonlinear airflow, a symplectic discretisation of the bore's 1D wave dynamics that aligns with known finite-difference time-domain results, and a new low-frequency effective model for tonehole impedance including a switching mechanism to simulate tonehole closure/opening. These elements are assembled such that the composite simulation remains passive and stable by construction. The benefit of the PHS approach is clearly demonstrated–once each component is cast in this form, they can be connected via power-conserving ports, and the overall instrument inherits guaranteed stability (no numerical energy gain) regardless of strong nonlinearity at the reed or rapid state changes at toneholes. Darabundit and Scavone validate their model on a virtual clarinet, showing that it can reproduce sustained oscillations and note transitions without instability.

2.3 Hardware-accelerated modal reverberation

Michon et al. tackle the question of real-time performance for large-scale physical models by comparing CPU, GPU, and FPGA implementations of a modal reverberation algorithm. Modal synthesis, which represents an object's vibration as a sum of many independent harmonic oscillators, is an attractive technique for artificial reverberation and resonator effects due to its physical interpretability and inherently parallel structure. However, simulating thousands of modes at audio rate is computationally intensive. This study provides a timely examination of how modern computing platforms fare in this task. The authors implement a high-order modal reverb (a plate reverb with thousands of modes) on three platforms: a multi-core CPU (with vectorisation and threading optimisations), a general-purpose GPU (using hundreds of parallel threads), and an FPGA (using custom hardware pipelines), each carefully optimised to exploit the architecture's strengths. They then measure maximum achievable modal complexity (number of modes), processing latency, and resource usage in various real-time scenarios. The results reveal a nuanced trade-off: GPUs excel at scalability, comfortably handling the largest number of modes due to their massive parallelism, whereas FPGAs deliver unparalleled low-latency processing, making them ideal for time-critical applications. Meanwhile, modern CPUs–benefiting from increasing core counts and SIMD vector units–show surprisingly strong performance for moderate polyphony, approaching the throughput of specialised hardware for mid-sized problems. These findings underscore that no single processor is "best" for all cases; instead, the choice depends on the specific requirements (e.g., maximum reverb length vs. latency tolerance vs. development flexibility). Michon et al. conclude with discussions on each platform's practical role in audio DSP–GPUs for heavy parallel workloads in studio or cloud settings, FPGAs for ultra-low-latency embedded systems, and CPUs for general-purpose use where moderate parallelism suffices.

2.4 Passive nonlinear string–bow interaction

Matusiak et al. present a rigorous study of the frictional interaction between a bowed string and the bow hair. Their work takes the elasto-plastic (E-P) friction law–valued for its ability to describe sticking, pre-sliding, and slip, but prone to spurious energy injection under naïve discretisation–and subjects it to detailed energy analysis. By reformulating the bristle damping term, they obtain a version of the model that is unconditionally passive, i.e., incapable of injecting net energy regardless of parameter

choice. Building on this, they derive a finite-difference scheme whose discrete energy mirrors the continuous balance, ensuring stability even in demanding transient regimes. A further theoretical advance, rarely achieved for friction models, is a proof of existence and uniqueness for the "bowed-mass" case. This result dispenses with the *ad hoc* remedies (caps on bow force, Friedlander's selection rules, or regularised Stribeck curves) often used to avoid Painlevé-type paradoxes and associated velocity jumps.

With this foundation in place, the authors embed the refined law into a complete bowed-string setting. Here, the bow is treated as a ribbon of finite width interacting with several adjacent string elements, its compliance represented explicitly, and the string's torsional motion is included alongside transverse vibration. These aspects, while known from earlier studies, are necessary for a credible rendering of how a bow distributes force and how torsion moderates slip behaviour. Numerical experiments span both lumped and distributed formulations, illustrating transients and steady Helmholtz motion. Where the original Dupont model may lose passivity or admit multiple solutions, the refined version remains stable, passive, and free of non-physical solution branches, reproducing measured slip amplitudes and spectral content without recourse to artificial constraints.

2.5 Port-Hamiltonian vocal synthesizer

Risse et al. extend physical modeling to voice synthesis with a reduced-order yet physically grounded model of the human vocal apparatus formulated in a port-Hamiltonian (pH) framework. A fluid–structure interaction governs the sound production mechanism: airflow from the lungs excites oscillations of the vocal folds, which modulate the flow into acoustic pressure waves in the vocal tract. Modeling voiced speech thus entails coupling a compressible airflow with vibrating tissue and a time-varying acoustic cavity. Risse et al. propose a quasi-1D distributed model for the glottal airflow and vocal tract, capturing effects such as cross-sectional area variation and air compressibility, and couple this with lumped representations of vocal fold dynamics and tract wall vibration. Each component is cast as a pH or energy-conserving subsystem, ensuring that when interconnected, power exchanges between flow, tissue, and acoustic radiation remain balanced, with no spurious energy gain or loss. A dedicated regularisation method is introduced to handle glottal closure (when the folds collide and airflow is momentarily cut off) in a numerically robust way. The continuous formulation is then discretised using structure-preserving methods, combining finite-volume schemes in space with an energy-consistent time integrator. This yields simulations that preserve the passivity of the full self-oscillating system while enabling real-time execution for the vocal tract. Numerical experiments explore both linear and nonlinear behaviour: frequency responses of static vocal-tract configurations (to verify formant placement), dynamic vowel transitions, and finally phonation, where the model generates self-sustained oscillations and synthesises vowel sequences with co-articulation effects. By combining theoretical rigour, computational efficiency, and stability, this work advances physically based voice synthesis beyond either unwieldy finite-element models or simplified low-order approximations prone to instability.

2.6 Differentiable modeling of nonlinear audio effects

Comunitá et al. examine how far differentiable learning can take the emulation of nonlinear audio effects. They map out the terrain of black-, grey-, and white-box strategies, then focus on the first two, comparing a wide set of architectures on guitar amplifiers, overdrive, distortion, fuzz, and compression. On the black-box side, they test LSTMs, temporal and gated convolutional networks, and structured state-space models; for grey-box, they propose block-oriented designs for compressors and for drive/fuzz circuits, using differentiable controllers to capture time-varying behaviour such as bias shift in fuzz pedals. A major practical contribution is the ToneTwist AFx dataset: forty analog and digital devices, recorded as dry/wet pairs over varied sources and playing styles, released with code and training scripts. Extensive experiments–objective metrics and listening tests–show no single method dominates across all devices, but highlight trade-offs: recurrent nets excel at strongly dynamic effects, convolutional nets at static nonlinearities. At the same time, grey-box models achieve good accuracy with far fewer parameters. The paper gives a clear picture of current capabilities and points to hybrid approaches and better training protocols as the next step for universal, differentiable models of audio effects.

3 Outlook and trends

The articles gathered here show how far physics-based sound synthesis has come, and how diverse its concerns have become. A first theme is the drive for physical accuracy with numerical robustness. Whether through port-Hamiltonian formulations or detailed treatments of nonlinear friction, the models respect energy or passivity constraints while remaining stable. This marks a shift from early virtual instruments, which often traded rigour for simplicity, toward systems that achieve both fidelity and reliability, helped by modern computing power.

A second thread is the use of machine learning as an ally rather than a rival: differentiable methods are being used to calibrate or extend physical models, or to learn sub-blocks (resonators, nonlinear mappings) under physical constraints. Virtual-analog, in particular, has seen a surge of black- and grey-box applications, marking a significant shift from traditional DSP techniques. Work on modal reverberation, such as large plate and room simulations, demonstrates how increased computing power through GPUs, FPGAs, and multicore CPUs has enabled the simulation of systems once out of reach.

There is also an apparent concern for real-time performance and playability. Many contributions focus on achieving low latency and efficient implementation, utilising model-order reduction, hardware parallelism, or explicit schemes where stability permits. Physical models become most compelling when they can be performed, and efficiency often sharpens rather than dilutes their essence.

Finally, it is worth noting that developments in physical modelling for audio and musical acoustics–where demands on efficiency, robustness, and sensitivity to time-varying parameters are unusually severe–often spill over into other domains. Methods pioneered in this community have found applications in control engineering, robotics, nonlinear systems, computational physics,

psychology, tribology, and signal processing. Examples include the adoption of nonlinear audio circuit simulation methods in robotics, the direct transfer of collision and friction modelling to robotic manipulation, the use of hardware-accelerated modal schemes for highly oscillatory systems, and the emergence of new tools for experimental psychology and musical practice.

Taken together, these papers underline the field's hybrid character: tools from computational mechanics, control, signal processing, and machine learning are being fused not only to advance the state of the art in physically modelled sound, but also to provide techniques and insights that resonate far beyond musical acoustics.

Author contributions

TH: Writing – review and editing. MVW: Writing – review and editing. MD: Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Michele Ducceschi's work was supported by the European Research Council (ERC) under the Horizon2020 programme, with grant NEMUS/950084.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note