

# State-space modeling of sound source directivity: An experimental study of the violin and the clarinet

Esteban Maestre, Gary P. Scavone, and Julius O. Smith

Citation: *The Journal of the Acoustical Society of America* **149**, 2768 (2021); doi: 10.1121/10.0004241

View online: <https://doi.org/10.1121/10.0004241>

View Table of Contents: <https://asa.scitation.org/toc/jas/149/4>

Published by the [Acoustical Society of America](#)

---

## ARTICLES YOU MAY BE INTERESTED IN

[Modelling and measurement of laser-generated focused ultrasound: Can interventional transducers achieve therapeutic effects?](#)

*The Journal of the Acoustical Society of America* **149**, 2732 (2021); <https://doi.org/10.1121/10.0004302>

[Substrate vibrations and their potential effects upon fishes and invertebrates](#)

*The Journal of the Acoustical Society of America* **149**, 2782 (2021); <https://doi.org/10.1121/10.0004773>

[Speech intelligibility in a realistic virtual sound environment](#)

*The Journal of the Acoustical Society of America* **149**, 2791 (2021); <https://doi.org/10.1121/10.0004779>

[Nonplanar metasurface for perfect absorption of sound waves](#)

*The Journal of the Acoustical Society of America* **149**, 2323 (2021); <https://doi.org/10.1121/10.0003435>

[Impact of non-individualised head related transfer functions on speech-in-noise performances within a synthesised virtual environment](#)

*The Journal of the Acoustical Society of America* **149**, 2573 (2021); <https://doi.org/10.1121/10.0004220>

[Mouse middle-ear forward and reverse acoustics](#)

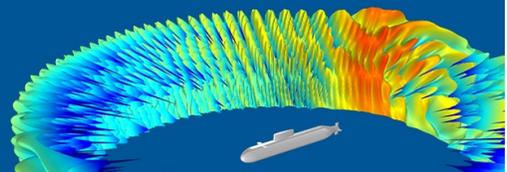
*The Journal of the Acoustical Society of America* **149**, 2711 (2021); <https://doi.org/10.1121/10.0004218>

---

**COMSOL Day**  
Acoustics

*A free, online event* where you can attend  
multiphysics simulation sessions, ask  
COMSOL staff your questions, and more

JOIN US MAY 25 »



## State-space modeling of sound source directivity: An experimental study of the violin and the clarinet<sup>a)</sup>

Esteban Maestre,<sup>1,b)</sup> Gary P. Scavone,<sup>1</sup> and Julius O. Smith<sup>2</sup>

<sup>1</sup>Computational Acoustic Modeling Laboratory, Schulich School of Music, McGill University, Montréal, Québec H2V4K2, Canada

<sup>2</sup>Center for Computer Research in Music and Acoustics, Music Department, Stanford University, Stanford, California 94350, USA

### ABSTRACT:

A method is presented for simulating the free-field, frequency-dependent directivity of linear sound sources for use in real-time within geometric acoustic environments. The method, which is applied to modeling the directivity of a violin body and a clarinet air column from experimental acoustic data in this study, is based on using minimum-phase measurements to design a state-space filter, allowing the interactive simulation of a time-varying number of radiated sound wavefronts, each toward a time-varying direction. With applicability in sound synthesis and/or auralization within virtual environments, where sound sources change position and orientation dynamically, techniques are proposed for modeling and simulating directivity profiles on perceptual frequency axes with alternatives for representing directivity on a per-vibration-mode basis while incorporating relative phase terms or by reduced-order efficient representations comprising separate components for the signature resonant structure and the associated directivity on an adjustable frequency resolution. © 2021 Acoustical Society of America.

<https://doi.org/10.1121/10.0004241>

(Received 29 September 2020; revised 14 March 2021; accepted 19 March 2021; published online 22 April 2021)

[Editor: Nicholas J. Giordano]

Pages: 2768–2781

### I. INTRODUCTION

Modeling and simulation of the frequency-dependent directivity of sound sources is a long-standing problem in acoustics with clear applications in computer animation, immersive media, and music synthesis. Given the role that music has played in the development of a number of basic principles for acoustics (Chaigne and Kergomard, 2016; Fletcher and Rossing, 1998), it is no surprise that musical instruments have often been the case study in recent research focusing on measurement and simulation of the sound radiation properties of sound sources.

After the pioneering work of Meyer (1972), several studies have proposed measurement techniques for characterizing the directivity of musical instruments in a performance situation, either for research on the properties of the instrument itself or with the purpose of room acoustic simulations or auralizations. Otondo and Rindel (2004) report a first comprehensive attempt to obtain a representation of the frequency-dependent directivity of a musical instrument during playing, using 13 microphones in an anechoic chamber at 45-deg intervals in the horizontal and vertical planes. The goal with this setup was not to acquire a detailed description of the whole radiation sphere of the instrument but to obtain a sample of the variations of the radiation (in seven octave bands) in both planes to be used afterward in room simulations. The method consisted of simultaneous

recordings of short isolated tones played on a clarinet, a trumpet, and a French horn with a similar musical intensity of mezzo forte over the whole performing pitch range, and asking the performers to hold the instrument without moving. A number of later works have improved on the angular resolution and included other instruments. Behler *et al.* (2012) and Pollow *et al.* (2009) used a 32-microphone array shaped as a truncated icosahedron to measure sustained tones of 41 symphonic orchestral instruments, which were held as still as possible by the musicians. These results were compiled into a database (Weinzierl *et al.*, 2017), which was subsequently analyzed by Shabtai *et al.* (2017) employing spectral-domain harmonic partial tracking to estimate the radiation efficiency in third-octave frequency bands. A similar study, albeit not based on the harmonic peak amplitude analysis but just on band energy averaging, was presented by Pätynen and Lokki (2010) for a similar set of instruments using an array of 22 microphones in pentagonal arrangement, including different dynamics and pitch heights per instrument. In a more recent study (Bodon, 2016), a capture system comprising an arc-shaped microphone array was used to gather high-resolution directivity measurements of musical instruments in a performance situation wherein the performer and the instrument were rotated to gather data for the full sphere surrounding the pair. Canclini *et al.* (2020) used multiple plenacoustic cameras and motion tracking of a violin to obtain radiation efficiency measurements.

After Weinreich (1982), others have attempted to measure instrument directivities by employing artificial excitation techniques. A first complete work, including measurements and real-time physical modeling synthesis,

<sup>a)</sup>This paper is part of a special issue on Modeling of Musical Instruments.

<sup>b)</sup>Electronic mail: esteban.maestre@mcgill.ca, ORCID: 0000-0002-0712-639X.

was presented in [Cook and Trueman \(1998\)](#), where six string instruments, including three guitars, a mandolin, and two fiddles, were excited by an impact hammer to obtain radiation responses by means of a dodecahedral microphone array. More recently, in a work by [Perez-Carrillo \*et al.\* \(2011\)](#), a musician performed bowed glissandi on a violin equipped with an electric pickup while simultaneously recording the pickup signal and the radiated sound via a microphone array surrounding the instrument; by means of a purposely devised deconvolution technique, data were then used for obtaining directional frequency responses, which could be used to auralize the pickup signal and obtain stereo sound from the violin while equipped with motion sensors. Similarly, [Shabtai \*et al.\* \(2015\)](#) artificially excited a violin by injecting an electric current to one of the strings under a magnetic field while recording the radiated sound. With regard to automatic excitation of wind instruments, [Grothe and Kob \(2013\)](#) employed an artificial player for blowing a bassoon to derive two-dimensional directivity patterns via continuous measurements in a turntable setup. Later, they extended their technique to high-resolution three-dimensional directivity measurements but instead used an impulsive sound source to excite the bassoon up to 1.5 kHz ([Grothe and Kob, 2019](#)). These works have provided data comprising per-band directivity diagrams or discrete directional responses but have not attempted to propose effective methods for simulating the directivity of moving sources in virtual environments based on volumetric or geometric acoustic representations.

Regarding source directivity simulation within volumetric acoustic frameworks, [Hacihabiboğlu \*et al.\* \(2007\)](#) incorporated frequency-independent directivity into digital waveguide three-dimensional meshes. With respect to finite-difference time-domain (FDTD) schemes, frequency-independent directivity diagrams were first used in [Escolano and López \(2007\)](#) for simulating sources in room acoustic simulations. More recently, [Bilbao \*et al.\* \(2019\)](#) included frequency-dependent directivity in the FDTD formulation of sources, leading to a compact representation involving a set of distinct finite impulse response digital filters with each applied to one of several source terms, each conveyed by a spherical harmonic (SH) basis function. These volumetric schemes, however, still remain prohibitive for practical application in interactive virtual environments due to their high computational cost. In that regard, a more attractive option is offered by geometric acoustic frameworks where the directional sound emission of moving sources can be more affordably incorporated while allowing free-field and reflected wavefronts to reach the listener. In an early example, [Huopaniemi \*et al.\* \(1994\)](#) used simplified source directivity models with fractional delay lines and recursive digital filters, each dedicated to a wavefront propagation path. A similar approach was recently presented in [Chaitanya \*et al.\* \(2020\)](#) as part of a hybrid system combining geometric acoustic simulation and convolution with pre-computed impulse responses where the authors devote one generic octave-band equalizer to each propagation path. A

frequency-domain technique involving precomputation via the boundary element method or equivalent source method (ESM) is proposed by [Mehra \*et al.\* \(2014\)](#), which used per-band directivity data to represent sources in terms of frequency-domain SH decompositions of low order; at run-time, these decompositions were combined with lookup tables holding precomputed pressure fields produced by elementary SH sources at regularly spaced positions within a given scene. This allowed the reconstruction (at rates between 10 or 15 Hz) of responses incorporating source directivity up to 1 kHz to be used within a convolution scheme.

As for sound synthesis frameworks incorporating source directivity, [James \*et al.\* \(2006\)](#) proposed a method for rendering sound from a modal representation obtained by finite element modeling (FEM) of computer-generated objects virtually excited by impulsive force signals. The method circumvents the need for solving the Helmholtz equation at run-time thanks to using ESM for precomputing an approximation in terms of a finite number of low-order multipole equivalent sources distributed across the geometry of each simulated object. This allows simulating the directional sound radiation via a superposition of a fixed number of damped exponentials (modes) for each of the equivalent multipole sources. In a subsequent work, [Wang and James \(2019\)](#) also obtain a modal approximation from FEM, but instead of using ESM, FDTD is used to evaluate the propagation and obtain a real-valued lookup table per mode; later, at run-time, tables provide a single amplitude per direction per mode, yielding an efficient approximation where direction-dependent phase terms are ignored.

This paper presents a method for simulating the free-field, frequency-dependent directivity of linear sound sources by designing a multi-output recursive digital filter from minimum-phase measurement data. The method can be used to construct efficient, yet, accurate sound synthesis and auralization models departing from the experimental data obtained from real-world testing or intensive numerical simulations. It is based on a state-space filter formulation allowing the simultaneous rendering of a time-varying number of radiated sound wavefronts for interactive geometric acoustic frameworks where sound sources dynamically change position and orientation. First, the off-line analysis of directivity data of a sound source involves estimating the eigenvalues and output matrix of a fixed-size, time-invariant state-space filter whose output signals correspond to sound pressure wavefronts radiated in a set of discrete static outward directions from the source. Then, regression and/or interpolation are used to enable the run-time generation of output matrices interactively, yielding the ability to obtain the sound pressure signals corresponding to a time-varying number of wavefronts radiated by a moving source, each toward a time-varying direction. The method, although of general applicability in simulating the sound source directivity within geometric acoustic frameworks, is demonstrated here to model a violin body and a clarinet air column from real-world data acquired via high-resolution radiation measurements.

Flexible techniques are proposed to simulate sources on a per-vibration-mode basis or by efficient representations comprising separate components for modeling the signature resonant structure and the associated directivity on an adjustable frequency resolution. In contrast to potential alternatives based on partitioned convolution with long impulse responses of a resonant nature, the method allows the time-varying operation of recursive filter structures designed over warped frequency axes, yielding low computational costs while maintaining accuracy in the lower frequency regions. As opposed to other recent modal frameworks for interactive sound synthesis of directional sound, the proposed method incorporates relative phase terms when combining modal contributions, thus, adding a degree of freedom for simulating directional magnitude responses from minimum-phase approximations.

The paper is structured as follows. First, Sec. II introduces the modeling framework, and Sec. III describes the measurement techniques employed to characterize the violin and the clarinet. Next, Sec. IV gives details on the methods proposed for obtaining efficient synthesis models, and Sec. V describes and discusses some obtained results. Finally, concluding remarks and pointers to future opportunities are provided in Sec. VI.

## II. MODELING FRAMEWORK

Our modeling framework is largely based on the method first introduced in [Maestre et al. \(2019\)](#) for simulation of acoustic sources and receivers. For completeness, the framework is first reintroduced as a general-purpose modeling tool for linear sound sources, and then it is adapted to the directivity study presented here.

### A. State-space formulation

The state-space approach, well understood and broadly used in several disciplines of physics, mathematics, and engineering ([Sparks, 1967](#)), is a powerful method that uses the theory of vector spaces and difference equations to represent dynamical systems through a flexible abstraction involving a recursive algebraic relationship between a vector  $\mathbf{x}$  of  $P$  input excitations, a vector  $\mathbf{s}$  of  $M$  linearly independent state variables, and a vector  $\mathbf{y}$  of  $Q$  output observations. In discrete-time, the state-space filter representation of a linear, time-invariant dynamical system can be written as

$$\begin{aligned} \mathbf{s}(n+1) &= \mathbf{A}\mathbf{s}(n) + \mathbf{B}\mathbf{x}(n), \\ \mathbf{y}(n) &= \mathbf{C}\mathbf{s}(n) + \mathbf{D}\mathbf{x}(n), \end{aligned} \quad (1)$$

where system matrix  $\mathbf{A}$  enables state variable recursion, matrix  $\mathbf{B}$  is formed by  $P$  row vectors with each mapping the  $p$ th input variable onto the  $M$  state variables, matrix  $\mathbf{C}$  is formed by  $Q$  column vectors with each mapping the  $M$  state variables onto the  $q$ th output variable, and matrix  $\mathbf{D}$  defines the feed-forward relationship between the inputs and outputs. Often, the matrix  $\mathbf{D}$  in state-space models of mechanical systems is defined to be zero as such systems

can be represented by strictly proper transfer functions. Virtually, an infinite number of valid representations will satisfy the input-output behavior observed for a particular dynamical system: given the inputs and outputs of the system, each possible choice of what the state variables represent will define a different *state-space* and imply a different algebraic relationship (i.e., a different set of system matrices) between the chosen state variables and vectors of inputs and outputs of the system. Among all valid representations, the *modal form* piques an interest: from a mathematics perspective, it involves a diagonal matrix  $\mathbf{A}$  and, thus, implies  $M$  decoupled state variables; from a physics perspective, the  $M$  state variables are proportional to the amplitudes of the  $M$  modes of the system. System identification techniques ([Ljung, 1999](#); [Söderström and Stoica, 1989](#)) enable designing a state-space filter from input-output measurements alone. For example, in structural mechanics, one may decide to choose a vector of surface driving forces as the input excitation vector and a vector of surface velocities as the output observation vector and design a filter to mimic the observed input-output behavior; if the designed filter is expressed in modal form, then the state variable vector will be proportional to the modal amplitudes of the system.

### B. Linear sound sources as time-varying state-space filters

In general, a complete identification of the motion of the continuous surface of a vibrating object is sufficient to describe acoustic sound radiation with sound propagating outward as pressure waves ([Williams, 1999](#)). For the purposes of directivity modeling, this work proposes to treat linear sound sources as dynamical systems substantially presenting minimum-phase behavior. Under this approximation, it is possible to employ the state-space abstraction to simulate the frequency-dependent directivity of a sound source by analyzing a set of input-output minimum-phase transfer functions obtained from measuring a finite number of radiated sound pressure signals in a finite number of discrete positions on the boundary of a domain enclosing the source. This allows simulating the directivity by deciding on a number  $M$  of state variables and estimating the system matrices from input-output measurements alone, irrespective of whether those measurements are obtained from testing a real sound source via acoustic measurements or virtual testing within a numerical simulation of any solid, fluid, or solid-fluid mechanics needed to reproduce the sound radiation properties of a mechanical structure.

As an example, let us assume a mechanical structure acting as a linear sound source for which (i) a continuous input domain is defined as the structure surface on which any number  $P$  of driving force signals, each at position  $\mathbf{u}^p$  of the surface, can be applied to excite the structure, and (ii) a continuous output domain is defined as an exterior enclosing sphere on which any number  $Q$  of radiated sound pressure signals, each corresponding to an angle  $\mathbf{v}^q$ , can be measured. Assuming no input-output feed-forward relationship, a modal-form, time-invariant state-space modeling abstraction

for the minimum-phase behavior of the sound source could require defining a dense set of  $P'$  discrete positions for the input domain and a dense set of  $Q'$  angles for the output domain, obtaining a  $P' \times Q'$  matrix of minimum-phase input-output impulse responses or transfer functions, deciding the number  $M$  of state variables, and accordingly estimating a diagonal transition matrix  $\mathbf{A}$ , an input matrix  $\mathbf{B}$  of  $P'$  rows with each corresponding to the  $p'$ th discrete excitation position, and an output matrix  $\mathbf{C}$  of  $Q'$  columns with each corresponding to the  $q'$ th discrete radiation angle, and imposing  $\mathbf{D} = 0$ .

However, as interactive geometric acoustic frameworks would typically entail the simulation of a direct-field wavefront and a time-varying number of early reflection wavefronts propagating from the mechanical structure to each listener, the number of radiated wavefronts and their angles would change over time as required to simulate how wavefronts radiated from a moving source get reflected from walls or obstacles before reaching a moving listener. The same goes for the input as both the number of driving forces and their positions on the surface may change over time: imagine how a rigid body simulation of the interaction between this structure and a solid environment (e.g., the structure falling and repeatedly bouncing off the ground) would lead to a time-varying number of contact (excitation) forces at time-varying positions of the structure surface. Instead of carrying out the simulation via a time-invariant state-space filter of the total  $P'$  inputs and  $Q'$  outputs, which would incur a high computational cost due to large input and output matrices and probably require purposely devised mechanisms to avoid discontinuity in inputs and outputs as excitation positions and radiation directions change over time, the state-space filter can be reformulated as a time-varying filter in which the system matrix  $\mathbf{A}$  is time-invariant but the sizes and coefficients of matrices  $\mathbf{B}$  and  $\mathbf{C}$  change over time as described below.

To accommodate for a time-varying number of inputs and/or outputs, let the row vectors of  $\mathbf{B}$  be referred to as *input mapping vectors*, and let the column vectors of  $\mathbf{C}$  be referred to as *output mapping vectors*. With this, the behavior of a sound source can be represented in terms of a mutable, time-varying state-space filter comprising input and/or output matrices of mutable size, respectively, presenting a time-varying number of input and/or output mapping vectors, each with time-varying coefficients. Here, the mutability and time-varying property of inputs and outputs does not convey that the modeled sound source object is changing over time; instead, it conveys that the position and/or orientation from where the sound source is excited and/or radiates sound is changing over time. In other words, the modeled dynamical system continues to be linear and time invariant, but mutable, time-varying input and output matrices are used to more elegantly simulate time-varying excitation and observation. In this context, without loss of generality, assuming no feed-forward input-output relationship and employing a more convenient vector notation, the discrete-time update relation of a sound source model is written as

$$\begin{aligned} \mathbf{s}(n+1) &= \mathbf{A}\mathbf{s}(n) + \sum_{p=1}^P \mathbf{b}^p(n)x^p(n), \\ y^q(n) &= \mathbf{c}^q(n)^T \mathbf{s}(n), \end{aligned} \tag{2}$$

where  $n$  is the time index,  $\mathbf{s}(n)$  is the vector of  $M$  state variables, matrix  $\mathbf{A}$  is diagonal and holds the vector  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m, \dots, \lambda_M]$  of  $M$  system eigenvalues along its diagonal,  $x^p(n)$  is the  $p$ th input (a scalar) of those existing at time  $n$ ,  $\mathbf{b}^p(n)$  is its corresponding length- $M$  vector of input mapping coefficients,  $y^q(n)$  is a  $q$ th system output (a scalar) obtained as a linear combination of the state variables, and  $\mathbf{c}^q(n)$  is the corresponding length- $M$  vector of output mapping coefficients. Note that the total number  $P$  of inputs is time varying, but the “ $(n)$ ” is dropped to simplify notation. The same goes for the total number  $Q$  of outputs, with  $q = 1, \dots, Q$  above. The  $p$ th product  $\mathbf{b}^p(n)x^p(n)$  is referred to as the  *$p$ th input mapping*, and the  $q$ th product  $\mathbf{c}^q(n)^T \mathbf{s}(n)$  is referred to as the  *$q$ th output mapping*. A similar filter could be formulated to include input-output feed-forward behavior in terms of analogously defined input-output mapping coefficients, but for reasons of simplicity, this is not considered in the remainder of this work. As it will be seen later, the physical units of the inputs of this filter are not restricted by definition (if coefficients are estimated from measurement data, input units will be proportional to the input excitation measurement units); the output, however, will likely be in units proportional to pressure (as commonly measured and simulated in geometric acoustic environments). It should be noted that, similar to classic state-space models with fixed-size, time-invariant matrices, this filter can also be realized via first- and/or second-order parallel sections (Bank, 2018), although it would present a time-varying number of inputs or outputs, with each involving time-varying.

Given a set of  $M$  eigenvalues, the time-varying behavior of the input and output mapping vectors is governed by respective *mapping functions*  $S^+$  and  $S^-$  taking time-varying coordinates associated with those inputs or outputs (e.g., the position at which the source object is excited or the outward direction of a radiated wavefront by the source) as arguments. This can be expressed as  $\mathbf{b}^p(n) = S^+(\mathbf{u}^p(n))$  where  $\mathbf{u}^p$  is a vector of coordinates in the input excitation domain for the  $p$ th input, and  $\mathbf{c}^q(n) = S^-(\mathbf{v}^q(n))$  where  $\mathbf{v}^q$  is a vector of coordinates in the output observation domain for the  $q$ th output. Mapping functions may be devised from arbitrary designs or, as outlined later in this work, from measurement data via regression models or simply interpolation.

### C. Our experimental study of source directivity

This work presents an experimental case study of simulating the directivity of a violin and clarinet as sound sources modeled by time-varying state-space filters similar to that of Eq. (2) but focused on the output-only behavior by imposing a fixed, time-invariant one-dimensional input excitation in each case: for the violin, the horizontal force on the bridge; for the clarinet, the fluid flow at the bore entrance

(after the reed channel). Note that the respective nonlinear excitation mechanisms are circumvented, treating the violin body and air column as linear systems. In both cases, the output corresponds to the free-field radiated pressure with the output domain coordinate space defining a sphere surrounding the instrument and determined by two angles,  $\theta$  and  $\varphi$ . Accordingly, in both cases, this leads to a state-space filter model of the form

$$\begin{aligned} \mathbf{s}(n+1) &= \mathbf{A}\mathbf{s}(n) + \mathbf{1}x(n), \\ y^q(n) &= \mathbf{c}^q(n)^T \mathbf{s}(n), \end{aligned} \quad (3)$$

where  $x(n)$  is the input excitation signal (in units proportional to  $N$  for the violin and  $\text{m}^3 \text{s}^{-1}$  for the clarinet),  $\mathbf{1}$  is a vector of ones (a design decision) and  $y^q(n)$  is the  $q$ th output signal corresponding to the sound pressure wavefront radiated toward a  $q$ th time-varying direction (in units proportional to Pa) with a vector  $\mathbf{c}^q(n)$  of time-varying coefficients provided by a mapping function  $S^-$  taking two spherical angles as arguments, i.e.,  $\mathbf{c}^q(n) = S^-(\theta^q(n), \varphi^q(n))$  with  $\theta^q$  and  $\varphi^q$ , respectively, referring to the azimuth and elevation of the  $q$ th radiated wavefront.

The choice of example sound sources used in the experimental study of directivity presented here, rather than restricting the applicability of the method, helps to demonstrate its flexibility: the observed sound radiation behavior of two physical sound sources of fairly different natures (a violin and a clarinet) can be simulated under the same framework by simply choosing the appropriate input excitation in each case, deciding a modeling order and estimating eigenvalues on a warped frequency scale, employing minimum-phase measurement data for estimating the output matrix of a time-invariant state-space filter, and using such a matrix for devising a regression-interpolation scheme to interactively provide direction-dependent output mapping vectors during real-time operation of a time-varying filter. As an application case for the source simulations studied in this paper, a real-time physical modeling synthesis for virtual reality is a clear example as both the bridge force in string instruments and the bore input flow in wind instruments can be obtained interactively from respective simulations (Maestre *et al.*, 2014; Maestre *et al.*, 2017a, 2018). Concerning the violin, a further application example was advanced by Maestre and Scavone (2019), who used a directivity model is used to augment the electrical pickup signal of a silent violin. Similar applications could instead involve processing of a contact microphone attached to the body of acoustic instruments (Rau *et al.*, 2018). In any case, given the principles of the well-known state-space system identification theory, this modeling technique could be similarly applied to simulate the linear part of any other sound source type (mechanical, electro-mechanical, or otherwise) as long as its input-output behavior is characterized by an appropriate set of impulse response or frequency response measurements obtained from real testing or from virtual testing within proper physics-based numerical simulations.

### III. MEASUREMENTS

Violin and clarinet radiation measurements were performed in a hemi-anechoic chamber with carpeted floor, resulting in extremely low reflectivity. The use of a fully anechoic chamber could potentially improve the quality of our results, but this limitation has no bearing on the applicability of the method. The sound radiation of a performance-level violin was measured by exciting the bridge with an impact hammer and measuring the sound pressure with a 64-microphone quasi-semi-circular array formed by 8 linear sub-arrays each comprising 8 pre-calibrated Sennheiser KE-4 capsules (Wedemark, Germany) as displayed in Fig. 1 (left), leading to approximately 7 cm of distance between capsules. The sampling rate was 48 kHz. A specially designed rig was used to hold the violin vertically on top of a pole attached to a rotating base with the strings damped by rubber bands as depicted in Fig. 1. The base was rotated by 5-deg increments for a total of 72 azimuth  $\theta$  spherical angles. For each  $i$ th rotation, at least 5 measurement repetitions, each providing 1 input force measurement and 60 usable sound pressure measurements (4 capsules from the top linear sub-array were discarded due to redundancy), were taken wherein the impact hammer was manually activated by way of a pendulum to which a thin thread was attached and pulled from a distance of 3 m [see Fig. 1 (right) for details]. With this, a set of 60 transfer functions

$$H_{\theta_i, \varphi_j}^v(\omega) = \frac{1}{K} \sum_{k=1}^K \frac{P_{k, \theta_i, \varphi_j}(\omega)}{F_{k, \theta_i}(\omega)} \quad (4)$$

were obtained for every  $i$ th rotation, where  $P_{k, \theta_i, \varphi_j}(\omega)$  is the frequency-domain version of the sound pressure measured by the microphone located at the  $j$ th inclination angle  $\varphi_j$  in the  $k$ th repetition, and  $F_{k, \theta_i}(\omega)$  is the frequency-domain version of the excitation force measured by the impact hammer in the  $k$ th repetition. From each  $i$ th set of at least 5 repetitions, each comprising non-averaged 60 transfer functions, 60 pairwise magnitude-square coherence functions were



FIG. 1. (Color online) (Left) Semi-circular 60-microphone array used for violin radiation measurements. The violin is held on the pole attached to the rotating base. (Right) Detail of the rig used for holding the violin, along with the arm to which the impact hammer pendulum is attached.

computed for each possible pair of repetitions; these functions were then integrated over frequency and summed across the 60 inclinations to get an aggregated pairwise coherence ranking, indicating the overall similarity between each pair of repetitions; then, the set of  $K = 3$  with the highest ranking were selected for averaging via Eq. (4). After distance-related gain compensation (the reader is reminded that the semi-circular array was formed by 8 linear sub-arrays) and conversion to minimum phase, this led to 4320 radiation transfer functions, corresponding to distinct positions on a centered spherical sector surrounding the instrument with a radius of 0.75 m from a chosen center, coinciding with the middle point between the bridge feet. The spherical sector being modeled covered over 90% of the sphere surface with  $-\pi < \theta < \pi$  and  $0 < \varphi < 60\pi/64$ . The angle  $\theta = 0$  corresponded to the direction perpendicular to the top plate, whereas the angle  $\varphi = 0$  corresponded to the direction pointed by the neck. Each measurement position corresponds to a pair  $(\theta_i, \varphi_j)$  of angles in the vertical polar convention, conforming to the output coordinates on a two-dimensional rectangular grid of  $60 \times 72 = 4320$  points. Such a grid corresponds to the uniform sampling of a two-dimensional Euclidean space whose dimensions are  $\theta$  and  $\varphi$ .

A similar setup was used to measure a performance-level soprano clarinet as depicted in Fig. 2 (left). The clarinet was excited with a driver attached to a six microphone probe calibrated with three nonresonant loads, measuring the bore input flow via a least-mean square signal processing technique as described in Lefevbre and Scavone (2011). The driver base was attached to the rotating base, and repetitions were analogously performed to obtain transfer functions,

$$H_{\theta_i, \varphi_j}^c(\omega) = \frac{1}{K} \sum_{k=1}^K \frac{P_{k, \theta_i, \varphi_j}(\omega)}{U_{k, \theta_i}(\omega)}, \quad (5)$$

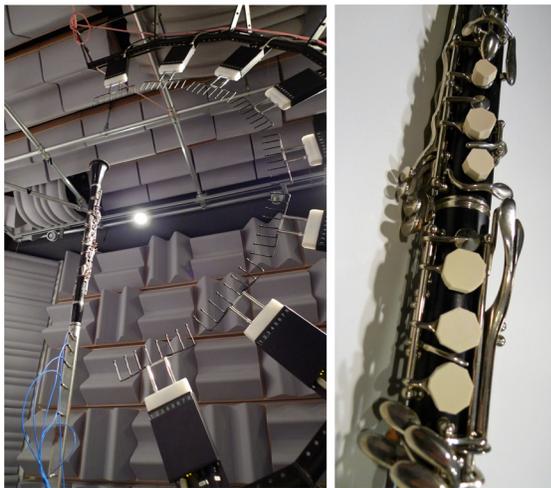


FIG. 2. (Color online) (Left) Semi-circular 60-microphone array used for clarinet radiation measurements. The clarinet exciter is held on the pole attached to the rotating base. (Right) Detail of the rubber plugs used for simulating the different fingerings.

where  $P_{k, \theta_i, \varphi_j}(\omega)$  is the frequency-domain version of the sound pressure measured by the microphone located at the  $j$ th inclination angle  $\varphi_j$  in the  $k$ th repetition, and  $U_{k, \theta_i}(\omega)$  is the frequency-domain version of the measured clarinet bore input flow in the  $k$ th repetition. The same steps were used to obtain 4320 transfer functions on a sphere with a chosen center coinciding with bell ring. The angle  $\theta = 0$  corresponded to the direction of the toneholes, whereas the angle  $\varphi = 0$  corresponded to the direction pointed to by the bell. To measure different fingerings, rubber plugs were cut to size and fitted to open toneholes as illustrated in Fig. 2 (right). The complete set of violin and clarinet measurements is available for download at <http://www.music.mcgill.ca/caml/doku.php?id=projects:direct> (Maestre *et al.*, 2018a).

#### IV. METHOD

The method for filter design and time-varying operation comprises three main steps: the estimation of eigenvalues is discussed in Sec. IV A, the estimation of the output matrix of a fixed-size, time-invariant state-space filter is reviewed in Sec. IV B, and spatial smoothing and resampling for time-varying, multi-wavefront synthesis operation is presented in Sec. IV C. As an efficient alternative, a method for the time-varying operation is described in Sec. IV D wherein a reduced-order state-space filter is designed to simulate the directivity of the instrument while conveying a prescribed vibration mode structure.

##### A. Eigenvalue estimation

The first step consists in defining or estimating the set of eigenvalues  $\lambda$  of Eq. (3). Recursive filters that simulate systems whose impulse responses are real-valued may present real eigenvalues and/or complex eigenvalues, with complex eigenvalues coming in complex-conjugate pairs. Although eigenvalues could be arbitrarily defined to tailor or constrain a desired behavior for the frequency response of the filter (e.g., by spreading eigenvalues over the complex disk to prescribe representative frequency bands), the eigenvalues corresponding to physical modes can be estimated from a set of target responses, which are representative of the behavior of the instrument. To do so, system identification techniques (Ljung, 1999; Söderström and Stoica, 1989) can be used to estimate a set of  $M$  eigenvalues. A simple procedure suitable for estimation over perceptual frequency resolutions is described next.

Given the decreasing frequency resolution in human hearing, convenient compromises between order  $M$  and perceptual quality can be reached through estimating eigenvalues on a perceptually motivated frequency scale. To do so, it is possible to use the discrete-time bilinear conformal map, which is defined by the all-pass substitution

$$z \leftarrow (\zeta + \rho)(1 + \rho\zeta)^{-1} \quad (6)$$

and takes the unit circle in the  $z$ -plane to the unit circle in the  $\zeta$ -plane in such a way that, for  $0 < \rho < 1$ , low

frequencies are stretched and high-frequencies are compressed as in a transformation from frequency in Hertz to a warped, perceptually motivated frequency scale (Härmä *et al.*, 2000; Smith and Abel, 1999). Warping of measured responses is performed by either applying the substitution of Eq. (6) to the impulse response measurements or attending to the phase response of the all-pass transfer function of Eq. (6), which leads to a relation between linear and warped frequency that can be realized by spline interpolation. Once the  $T = 4320$  measured responses have been mapped, autoregressive modeling (Bank and Karjalainen, 2010) can be used to estimate  $M$  eigenvalues as the roots of a polynomial  $\mathbf{p} = [1 \ \mathbf{a}]$ , with  $\mathbf{a} = [a_1, \dots, a_m, \dots, a_M]$  obtained by solving the weighted minimization

$$\underset{\mathbf{a}}{\text{minimize}} \sum_{t=1}^T w_t \sum_{l=M}^L \left( h'_t(l) + \sum_{m=1}^M a_m h'_t(l-m) \right)^2, \quad (7)$$

where  $h'_t$  are the warped-frequency impulse response measurements,  $w_t$  are the scalar weights applied to the responses, and  $L \geq M$  is the number of samples used per response. Given the uneven distribution of measurement positions on the sphere, each  $w_t$  is defined to be inversely proportional to the area of the respective  $t$ th polygon obtained after Voronoi tessellation (Voronoi, 1908). Once  $M$  warped-frequency eigenvalues are estimated, their linear-frequency counterparts are obtained back from Eq. (6).

Particularly in the lower frequencies and thanks to frequency warping, eigenvalues appearing in complex-conjugate pairs will often provide a reliable estimation for perceptually relevant resonant modes of the modeled source. If only vibration modes are to be considered in the model, real-valued eigenvalues can be discarded. Alternatively, additional immittance measurements may instead be used to estimate complex eigenvalue pairs via nonlinear optimization techniques as described by Maestre *et al.* (2013, 2017b).

### B. Estimation of output mapping coefficients

The second step consists of using measurement data and the  $M$  eigenvalues to estimate the output matrix  $\mathbf{C}$  of a fixed-size, time-invariant state-space filter akin to that in Eq. (1), whose transfer function is  $\mathbf{H}(z) = \mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$ . In multiple-input, multiple-output state-space systems, one would normally need to jointly estimate input and output matrices. In the case study of this paper, however,  $\mathbf{D} = 0$  and  $\mathbf{B} = \mathbf{1}$  [see Eq. (3)], and the transfer function reduces to  $\mathbf{H}(z) = \mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{1}$ . This means that it is possible to estimate the coefficients of  $\mathbf{C}$  by solving  $T$  minimization problems (each corresponding to a measurement angle), each in terms of the responses of  $M$  basis transfer functions  $H_m(z) = z^{-1}(1 - \lambda_m z^{-1})^{-1}$ . Assuming  $R$  real eigenvalues  $\lambda_r$  and  $J$  pairs of complex-conjugate eigenvalues  $(\lambda_j, \lambda_j^*)$  with  $M = R + 2J$  and imposing complex-conjugate symmetry to ensure realness, the problem

$$\underset{\mathbf{c}_t}{\text{minimize}} \sum_{l=0}^{L-1} \left( h_t(l) - \sum_{r=1}^R c_r h_r(l) - 2\Re \sum_{j=1}^J c_j h_j(l) \right)^2 \quad (8)$$

is solved for each  $t$ th angle, where  $\mathbf{c}_t$  is the  $t$ th row of matrix  $\mathbf{C}$  and comprises  $R$  real-valued coefficients  $c_r$  and  $J$  complex-conjugate pairs of coefficients  $(c_j, c_j^*)$ ,  $h_t$  is the response measurement,  $h_r$  are the responses of the  $R$  basis transfer functions corresponding to real eigenvalues  $\lambda_r$ ,  $h_j$  are the responses of the  $J$  basis transfer functions corresponding to complex eigenvalues  $\lambda_j$ , and  $L \geq M$  is the number of samples used per response. Note that the coefficients of matrix  $\mathbf{C}$  can be equivalently estimated over a perceptually motivated warped frequency axis, employing a method similar to what is described by Maestre *et al.* (2016) for frequency-domain responses.

### C. Spatial processing for run-time operation

At run-time, the vector of mapping coefficients  $\mathbf{c}^q(n)$  of Eq. (3) needs to be obtained interactively for each  $q$ th simulated wavefront as a function of the corresponding angle, i.e.,  $\mathbf{c}^q(n) = S^-(\theta^q(n), \varphi^q(n))$ . To allow this, in an off-line spatial pre-processing step, the estimated matrix  $\mathbf{C}$  is used to devise the function  $S^-$ . Notwithstanding the possibility of devising  $S^-$  as a parametric regression in terms of elementary basis functions (e.g., SHs), interpolation of known coefficient vectors stored in lookup tables remain a cost-effective solution. Still, it is advantageous to use a regression model for noise reduction, spatial smoothing, or resampling to generate the tables used for run-time interpolation. As an example, it is possible to obtain a SH model of the estimated mapping coefficients appearing in  $\mathbf{C}$  and then evaluate the model to populate a lookup table of a desired size wherein the distribution of coefficients is more continuous and less prone to artifacts during run-time interpolation. Let  $\mathbf{c}_m$  be the  $m$ th column vector of the estimated matrix  $\mathbf{C}$ , associated with the  $m$ th state variable and comprising  $T$  coefficients, with each  $t$ th coefficient corresponding to the direction defined by the angle  $(\theta_t, \varphi_t)$  in the original set of  $T = 4320$  measurements. The  $\mu$ th-order SH basis functions (Zotter, 2009) are

$$Y_\mu^\eta(\theta_t, \varphi_t) = k_\mu^\eta e^{im\varphi_t} P_\mu^\eta \cos(\theta_t), \quad \mu \in \mathbb{N}, \quad -\mu \leq \eta \leq \mu,$$

where  $P_\mu^\eta$  are the associated Legendre polynomials and  $k_\mu^\eta$  are the normalization factors

$$k_\mu^\eta = \left( ((2\mu + 1)(\mu - \eta)!) (4\pi(\mu + \eta)!)^{-1} \right)^{1/2}.$$

By constructing a matrix  $\mathbf{Y}_\mu$  of size  $(\mu + 1)^2 \times T$  containing the basis functions up to order  $\mu$  and evaluated for all  $T$  measurement angles, the weighted minimization

$$\underset{\mathbf{g}_m}{\text{minimize}} \|\mathbf{W}^{1/2}(\mathbf{c}_m - \mathbf{Y}_\mu \mathbf{g}_m)\|^2$$

can be performed where  $\mathbf{W}$  is a diagonal matrix of weights  $w_t$  procured, again, from Voronoi tessellation to obtain the performed where  $\mathbf{W}$  is a diagonal matrix of weights  $w_t$  procured again from Voronoi tessellation, and obtain the regression coefficients  $\mathbf{g}_m$ , associated with the  $m$ th state variable. At run-time, given  $\mathbf{g}_m$  and an angle  $(\theta_q, \varphi_q)$ , it would be straightforward to obtain  $Y_\mu^q(\theta_q, \varphi_q)$  and then compute  $c_m^q$  by linear projection. However, as mentioned above, it is more convenient to instead use the regression model for precomputing lookup tables to be used for interpolation at run-time.

**D. Reduced-order modeling**

From inspecting Eq. (3), it can be deduced that the computational complexity of the filter will be proportional to  $(M + MQ)$ , where  $M$  is the order of the filter,  $Q$  is the number of simulated wavefronts, and the cost of output mapping is proportional to  $(MQ)$  and becomes dominant as  $Q$  increases. The order  $M$  can be arbitrarily set to achieve a desired compromise between the modeling fidelity and the computational cost associated with each simulated wavefront. However, and especially in the context of sound synthesis applications, the lowering of  $M$  may lead to the loss of the signature modal structure of the sound source. This is particularly relevant in the case study of this work in which vibration modes determine not only the directivity but also the signature sound. To cope with this problem while still achieving a reduction of the overall computational cost in multi-wavefront simulation, a strategy for reduced-order modeling is proposed as follows.

Let  $\lambda$  be a set of  $M = 2J$  eigenvalues in a model comprising  $J$  complex-conjugate pairs and no real eigenvalues, able to faithfully represent the signature modal structure of a vibrating sound source. Let  $\omega_j$  be the natural frequencies, corresponding to each  $j$ th eigenvalue pair (i.e., each  $j$ th mode), and let  $\tilde{\Xi}(\omega)$  be a reference radiation transfer function for the sound source, chosen arbitrarily from the measurement set or obtained by minimum-phase conversion of a (weighted) average of all  $T$  measurements. From  $\tilde{\Xi}(\omega)$ , the output mapping vector  $\mathbf{c}$  of a single-output, time-invariant filter

$$\begin{aligned} \mathbf{s}(n + 1) &= \mathbf{A}\mathbf{s}(n) + \mathbf{1}x(n), \\ u(n) &= \mathbf{c}^T\mathbf{s}(n) \end{aligned} \tag{9}$$

is designed via least squares, similarly as described in Sec. IV B. By attending to the modeled reference response  $\Xi(\omega)$ , it is then obtained for each  $t$ th measurement  $H_t(\omega)$ , a set of  $J$  magnitude ratios  $|H_t(\omega_j)|/|\Xi(\omega_j)|$ . From these  $JT$  ratios,  $T$  minimum-phase target responses  $\hat{h}_{\xi,t}$  are then reconstructed via interpolation over the frequency axis and conversion to the time domain. The responses  $\hat{h}_{\xi,t}$ , each corresponding to a measurement angle  $(\theta_t, \varphi_t)$ , are then used to design a time-varying state-space filter

$$\begin{aligned} \mathbf{s}_\xi(n + 1) &= \mathbf{A}_\xi\mathbf{s}_\xi(n) + \mathbf{1}u(n), \\ y^q(n) &= \mathbf{c}_\xi^q(n)^T\mathbf{s}_\xi(n) \end{aligned} \tag{10}$$

of reduced order  $M_\xi < M$  by again following the steps described in Secs. IV A–IV C. In this configuration for which the computational complexity of output mapping becomes proportional to  $(M_\xi Q)$ , the time-invariant, single output filter of Eq. (9) conveys the signature modal structure of the sound source, whereas the reduced-order, mutable state-space filter of Eq. (10) models directivity at a coarser level and allows the simultaneous simulation of multiple outgoing wavefronts toward time-varying directions at a reduced cost per wavefront. This becomes particularly useful in interactive virtual environments where the simulation of multiple sound radiation paths may be needed (e.g., multiple direct-field and reflection signals), making it possible to reduce the computational requirements devoted to directivity modeling while preserving the signature vibration mode structure of the sound source.

**V. RESULTS**

This section first provides examples of directivity models obtained in a per-vibration-mode basis as described in Secs. IV A–IV C. Then, examples illustrating the reduced-order modeling approach described in Sec. IV D are presented.

**A. Per-mode directivity**

**1. Violin**

For the violin, data are first used to estimate eigenvalue sets over a perceptual frequency axis (bilinear conformal map,  $\rho = 0.85$ ) as described in Sec. IV A, leading to modeling orders  $J = M/2 = 8, 19, 38$  and comprising no real-valued eigenvalues. To illustrate the ability of the eigenvalue sets to convey the signature vibration mode structure of the violin, Fig. 3 displays the frequency response (200–6000 Hz) and impulse response (the first 20 ms) models obtained for the direction  $\theta = 0.95\pi, \varphi = 1.84\pi$ , correspondingly superimposed on the measured frequency and impulse response, respectively, over logarithmic frequency and time axes for a better illustration of the modeling accuracy. As it can be observed, using a warped frequency axis allows us to retain detail in the lower frequency region while yielding a smoother approximation over higher frequencies. Higher order models provide increasingly better approximations over a wider frequency band and a closer match in the time domain.

With regard to SH decomposition and lookup table resynthesis described in Sec. IV C, a model with  $J = 38$  eigenvalue pairs is used to depict, in Fig. 4, the distribution of the output mapping coefficient corresponding to the positive-imaginary eigenvalue of a pair representing a vibration mode presenting a low-complexity radiation pattern, with natural frequency near 1740 Hz over the two-dimensional Euclidean space determined by the original  $72 \times 60$  measurement angles (left plots) and a resynthesized table (right plots) with  $64 \times 64$  angles after SH decomposition of order  $\mu = 16$ . The distribution observed in Fig. 4 for

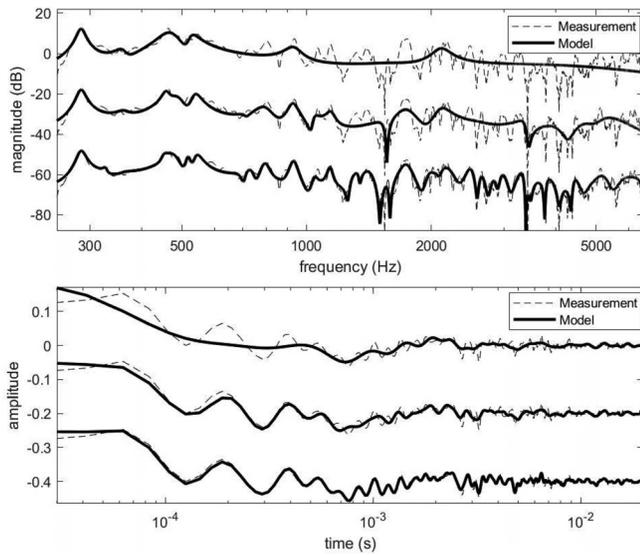


FIG. 3. Comparison of measured and modeled responses for the violin for the direction  $\theta = 0.95\pi$ ,  $\varphi = 1.84\pi$  and different modeling orders. (Top) Frequency responses of orders  $J = 8, 19, 38$ , top to bottom, respectively, offset for clarity. (Bottom) Impulse responses of orders  $J = 8, 19, 38$ , top to bottom, respectively, offset for clarity.

the modeled vibration mode suggests the sound radiation by the region of the  $f$ -holes of the violin (Bissinger *et al.*, 2007). Regarding the vertical lines appearing in the left plots in Fig. 4, it is assumed that they represent measurement noise in the form of slight amplitude variations across different angular positions of the rotating base, leading to fictitious directivity artifacts of spatial frequencies higher than those of the vibration modes of interest; such noise can be eliminated by SH order truncation as illustrated by the example.

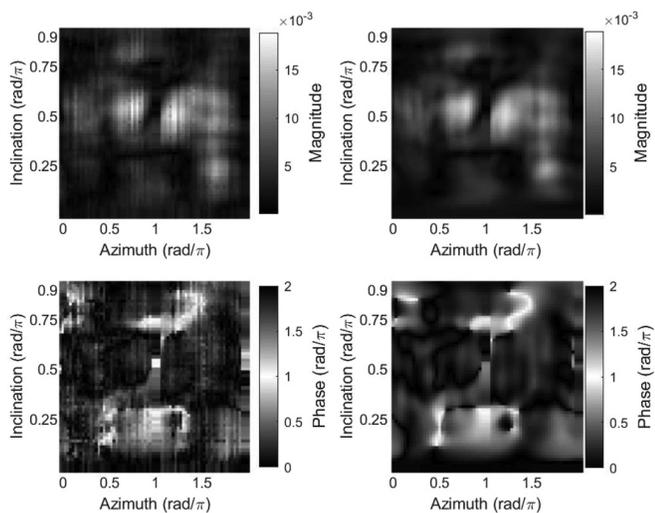


FIG. 4. (Left) Output mapping coefficient distributions over the sphere, corresponding to the positive-imaginary eigenvalue of a pair representing a vibration mode with natural frequency near 1740 Hz. (Top-left) Magnitude, estimation from measurements; (top-right) magnitude, SH model,  $\mu = 16$ ; (bottom-left) phase, estimation from measurements; and (bottom-right) phase, SH model,  $\mu = 16$ .

Finally, the model with  $J = 38$  modes is also employed to demonstrate the time-varying operation. For 512 consecutive steps, the output coordinates (angles  $\theta$  and  $\varphi$ ) of an outgoing wavefront as captured by an ideal microphone lying on the sphere surrounding the violin are slewed, simulating a continuous motion from the initial orientation at  $(\theta = 0, \varphi = 0.1\pi)$  to a final orientation at  $(\theta = 2\pi\text{rad}, \varphi = 0.8\pi)$ . Figure 5 compares the measured responses (nearest neighbor) and the modeled responses as obtained from the bilinear interpolation of output mapping coefficients in the resynthesized table. For reference, Fig. 5 also includes the response corresponding to an amplitude-only model obtained by constraining coefficients  $c_j$  in Eq. (8) to be real and performing bilinear interpolation of real-valued tables, similarly to the simplification employed in the method presented by Wang and James (2019) where direction-dependent phase is discarded and modes are combined with real-valued weights. As it can be observed, the additional degree of freedom offered by allowing relative phase terms in mode contributions yields a much better approximation across the frequency spectrum.

## 2. Clarinet

As with the violin, sets of complex-conjugate pairs of eigenvalues (bilinear conformal map,  $\rho = 0.85$ ) are first estimated, leading to modeling orders  $J = M/2 = 8, 17, 31$  for the fingering G3 (196 Hz), and orders  $J = 8, 16, 33$  for

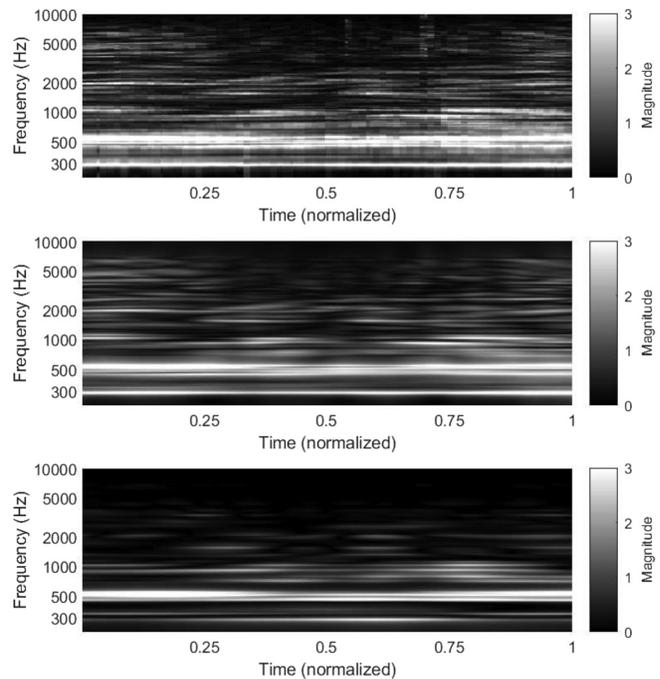


FIG. 5. Time-varying operation of a violin directivity model, simulating a linear trajectory on the space of orientation angles over 512 time steps from  $(\theta = 0, \varphi = 0.1\pi)$  to  $(\theta = 2\pi, \varphi = 0.8\pi)$ . (Top) Original frequency response obtained by nearest neighbor. (Middle) State-space filter model with  $J = 38$  modes via bilinear interpolation over a table resynthesized by SH decomposition. (Bottom) Amplitude-only model with  $J = 38$  modes, similar to Wang and James (2019), via bilinear interpolation over a real-valued table resynthesized by SH decomposition.

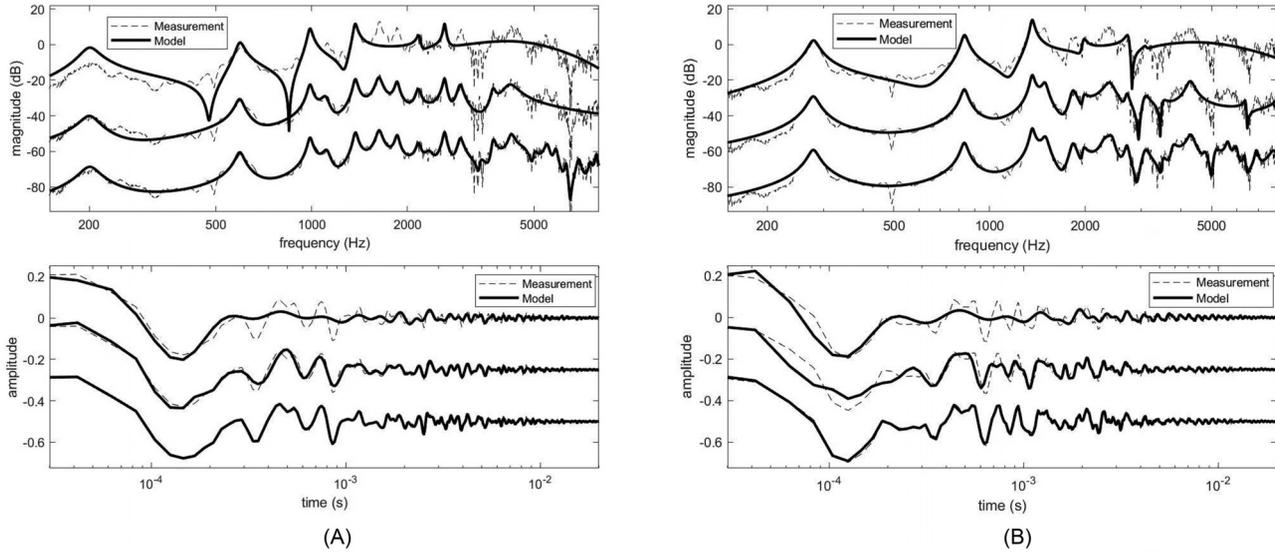


FIG. 6. Comparison of measured and modeled responses of the clarinet for the direction  $\theta = 0.38\pi$ ,  $\varphi = 0.55\pi$ . (A) G3 fingering, (top) frequency responses of orders  $J = 8, 17, 31$ , top to bottom, respectively, offset for clarity; (bottom) corresponding impulse responses, top to bottom, respectively, offset for clarity. (B) C4 fingering, (top) frequency responses of orders  $J = 8, 16, 33$ , top to bottom, respectively, offset for clarity; (bottom) corresponding impulse responses, top to bottom, respectively, offset for clarity.

the fingering C4 (262 Hz). Figure 6 displays the frequency response (150–8000 Hz) and impulse response (the first 20 ms) models obtained for the direction  $\theta = 0.95\pi$ ,  $\varphi = 1.84\pi$ , superimposed on the measurements. Again, using a warped frequency axis allows us to retain detail in the lower frequency region while yielding a smoother approximation over higher frequencies. In particular, for the fingering G3, the model of order  $J = 8$  seems insufficient to properly approximate the band around the first vibration mode; this problem, however, could perhaps be addressed by increasing the warping parameter  $\rho$  at the cost of achieving a worse approximation at higher frequencies. Again,

warped frequency designs allow us to retain detail in the lower frequency region in spite of moderately low orders. Higher orders yield better approximations over a wider frequency band and a closer match in the time domain.

The same process was used for SH decomposition and lookup table resynthesis as was used with the violin. The fingerings G3 and C4 and respective models with  $J = 31$  and  $J = 33$  eigenvalue pairs are used to depict, in Fig. 7, the distribution of the output mapping coefficient corresponding to the positive-imaginary eigenvalue of a pair representing the fifth vibration mode in each case with natural frequencies near 1340 Hz and 1810 Hz, respectively. The

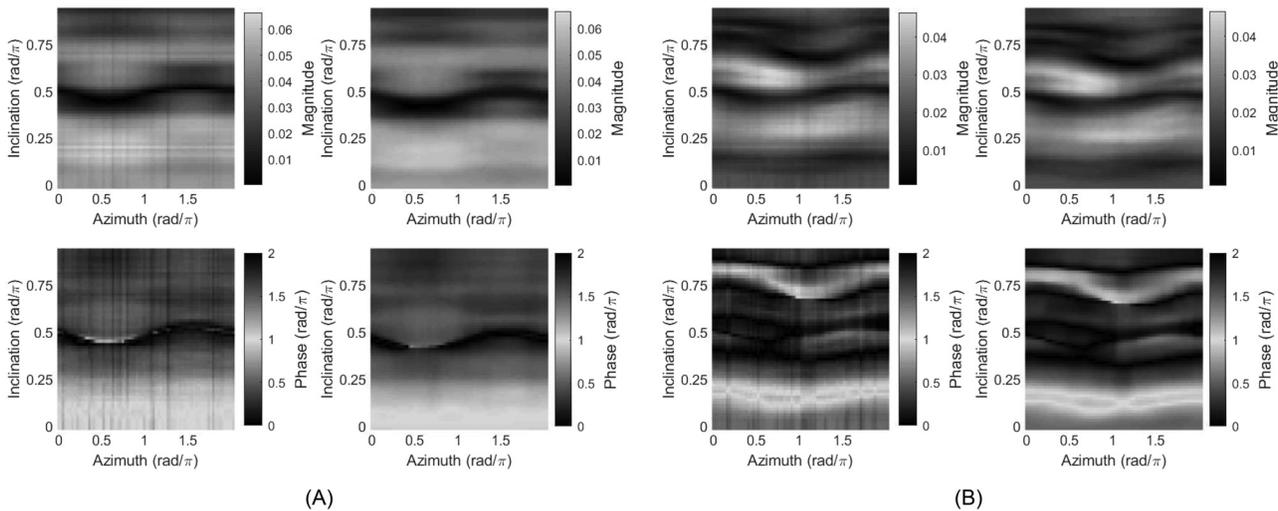


FIG. 7. Output mapping coefficient distributions over the sphere, corresponding to the positive-imaginary eigenvalue of a pair representing the fifth vibration mode of the clarinet. (A) G3 fingering,  $J = 31$  modes, the fifth mode is at natural frequency near 1370 Hz. (B) C4 fingering,  $J = 33$  modes, the fifth mode is at natural frequency near 1830 Hz. [(A),(B)] (Top-left) Magnitude, estimation from measurements; (top-right) magnitude, SH model,  $\mu = 20$ ; (bottom-left) phase, estimation from measurements; (bottom-right) phase, SH model,  $\mu = 20$ .

distributions are over the same two-dimensional Euclidean spaces as for the violin after SH decomposition of order  $\mu = 20$ . It is straightforward to notice a higher complexity in the pattern displayed by the C4 fingering, which involves three open toneholes as compared with the G3 fingering (no open toneholes). As it happened with the violin, measurement noise appears in the form of vertical lines, which can be removed by spatial smoothing or SH order truncation.

Finally, models with orders  $J = 31$  and  $J = 33$  are used again to demonstrate the time-varying operation for the G3 and C4 fingerings, respectively. As with the violin, the continuous linear motion of an ideal microphone on the sphere is simulated, this time following two different trajectories: first, from the initial position at  $(\theta = -\pi, \varphi = 0.4\pi)$  to a final position at  $(\theta = \pi, \varphi = 0.4\pi)$  along the azimuth axis and, then, from the initial position at  $(\theta = 0, \varphi = 0.1\pi)$  to a final position at  $(\theta = 0, \varphi = 0.8\pi)$  along the inclination axis. For each fingering and trajectory, Fig. 8 compares the measured responses (nearest neighbor) and modeled responses as obtained from the bilinear interpolation of output mapping coefficients in the resynthesized table. Again, for reference, Fig. 8 also includes the radiation responses obtained by an amplitude-only model (bilinear interpolation), similar to the approximation used in the synthesis model of Wang and James (2019). In the cases of Figs. 8(A) and 8(C), which correspond to the G3 fingering, a slight overestimation of the amplitude of the first mode can be observed; in the case of Fig. 8(C), the overestimation is accompanied by a slight modulation along the azimuth axis. With respect to the slight overestimation, it is attributed to the remarkably low amplitude of the first mode as compared to the modes above, which in the measurements [see Fig. 6(A)] can be easily observed as a less clearly defined resonance, significantly contaminated by noise—the reader is reminded here that the chamber in which the measurements were taken, although of low reflectivity, was not completely anechoic. With respect

to the modulation observed for the amplitude of the first mode in Fig. 8(C), a potentially straightforward solution could be to selectively lower the SH truncation order when dealing with lower frequency eigenvalues.

## B. Reduced-order directivity

### 1. Violin

Following the method described in Sec. IVD,  $J = M/2 = 39$  eigenvalue pairs are used to obtain three reduced-order models for the violin, each comprising both real- and complex-valued eigenvalues with  $M_\xi = 24, 16, 8$ . To obtain the reference time-invariant model, a weighted average of all  $T$  measurements is first performed with weights obtained by Voronoi tessellation. The reduced-order model is then built by SH modeling ( $\mu = 16$ ) and resynthesis as previously described.

To illustrate the behavior of the reduced-order models, the time-varying operation is demonstrated as with the full-order model of Sec. VA1, simulating the linear trajectory of an ideal microphone lying on the measurement sphere on the Euclidean space of orientation angles over 512 time steps from the initial position at  $(\theta = 0, \varphi = 0.1\pi)$  to a final position at  $(\theta = 2\pi, \varphi = 0.8\pi)$ . Figure 9 compares the original frequency response measurement (nearest neighbor) and reduced-order models (bilinear interpolation). For the case  $M_\xi = 24$ , it is possible to observe how the computational cost per outgoing wavefront can be reduced by an approximate factor of 3.17 while still preserving much of the detail, whereas the case  $M_\xi = 8$  (reduction by a factor of 9.5) yields a less accurate model of directivity. In both cases, however, the signature modal structure is maintained thanks to the time-invariant reference model. Selecting the appropriate orders for the reference and reduced-order model will depend on the availability of real-time computing power

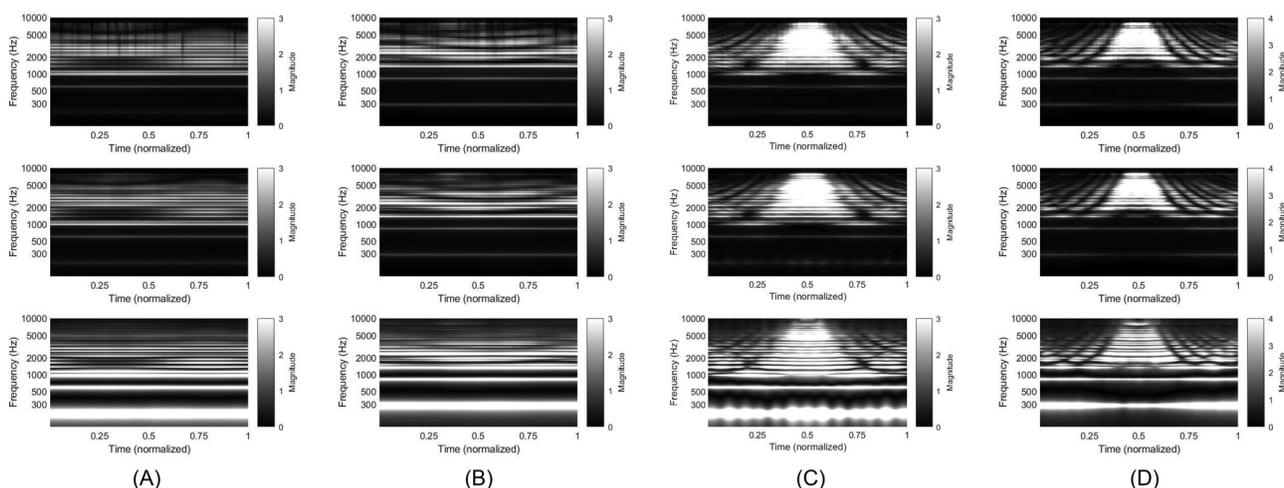


FIG. 8. Time-varying operation of clarinet directivity models. [(A),(C)] G3 fingering model with  $J = 31$  vibration modes and [(B),(D)] C4 fingering model with  $J = 33$  vibration modes. For all cases, a linear trajectory on the space of the orientation angles is simulated over 512 time steps: [(A),(B)] from  $(\theta = -\pi, \varphi = 0.4\pi)$  to  $(\theta = \pi, \varphi = 0.4\pi)$  and [(C),(D)] from  $(\theta = 0, \varphi = 0.1\pi)$  to  $(\theta = 0, \varphi = 0.8\pi)$ . [(A),(B),(C),(D)] (Top) Original frequency response, obtained by nearest neighbor; (middle) state-space filter model via bilinear interpolation over a table resynthesized by SH decomposition; (bottom) amplitude-only model, similar to Wang and James (2019), via bilinear interpolation over a table resynthesized by SH decomposition.

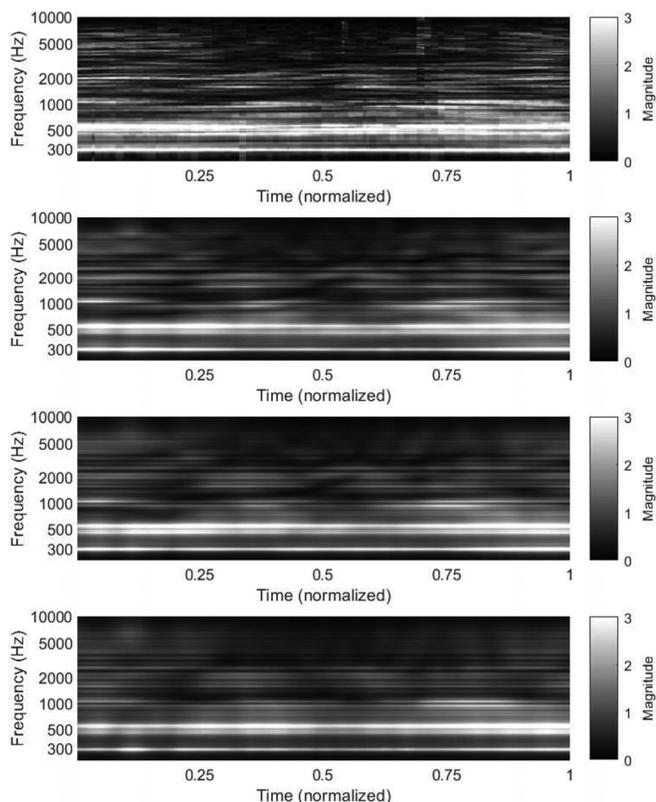


FIG. 9. Time-varying operation of a violin directivity model of reduced order via a reference model with  $J = 38$  vibration modes ( $M = 76$ ), simulating a linear trajectory on the space of orientation angles over 512 time steps from  $(\theta = 0, \varphi = 0.1\pi)$  to  $(\theta = 2\pi, \varphi = 0.8\pi)$ . (Top) Original frequency response, obtained by nearest neighbor; (second) reduced-order model of order  $M_\xi = 24$  via bilinear interpolation over a table resynthesized by SH decomposition; (third) reduced-order model of order  $M_\xi = 16$  via bilinear interpolation over a table resynthesized by SH decomposition; and (bottom) reduced-order model of order  $M_\xi = 8$  via bilinear interpolation over a table resynthesized by SH decomposition.

while accounting for the advantages brought forward by frequency warping.

## 2. Clarinet

For the clarinet, two sets of  $J = M/2 = 31$  and  $J = 33$  eigenvalue pairs are used, respectively, corresponding to the G3 and C4 fingerings, to obtain three reduced-order models, each comprising both real- and complex-valued eigenvalues, again, with  $M_\xi = 24, 16, 8$ . The reference time-invariant model is obtained again as a weighted average of all  $T$  original measurements. Then, the reduced-order model is built again by SH modeling ( $\mu = 20$ ) and resynthesis.

The time-varying operation of the reduced-order models is demonstrated similarly as with the violin over 512 time steps from the initial orientation at  $(\theta = 0, \varphi = 0.1\pi)$  to a final orientation at  $(\theta = 2\pi, \varphi = 0.8\pi)$  for each fingering model. Figure 10, again, compares the original frequency response measurement (nearest neighbor) and reduced-order models (bilinear interpolation). Congruent findings are observed. For the case  $M_\xi = 24$ , which allows a computational cost reduction by an approximate factor of

2.58 and 2.75, fairly satisfactory approximations are obtained. The case  $M_\xi = 8$  (reduction by a factor of 7.75 and 8.25) yields a less accurate directivity representation. Again, the signature modal structure is maintained thanks to the time-invariant model. Choosing the most appropriate order for both the reference model and reduced-order model will be a matter of compromise between cost and fidelity in light of an informed selection of the amount of frequency warping.

## VI. CONCLUSION AND FUTURE WORK

A method for efficiently simulating the sound source directivity in interactive geometric acoustic frameworks has been presented. With applications both in sound synthesis and spatial audio, the method provides a flexible scheme to simulate sources on a per-vibration-mode basis or by efficient representations comprising separate components for modeling the signature resonant structure and the associated directivity on an adjustable frequency resolution. In contrast to potential alternatives that are based on partitioned convolution with long impulse responses of a resonant nature, the method allows the efficient simulation of a time-varying number of radiated wavefronts by operating on a time-varying state-space filter designed over warped frequency axes, yielding low computational costs while maintaining accuracy in the lower frequency regions. As opposed to other recent modal frameworks for interactive sound synthesis of directional sound, the proposed method incorporates relative phase terms when combining modal contributions, thus, adding a degree of freedom for simulating the directional magnitude responses from the minimum-phase approximations. Rather than performing a detailed study of the violin and clarinet directivities, these instruments were chosen as challenging real-world examples of a diverse nature to showcase the capabilities of a method that can be applied to construct simulation models either from real experimental data or numerically generated data.

Depending on the application case and computational constraints, an optimal trade-off between cost and perceptual quality could be attained via subjective tests, and those are left for future studies. It should be straightforward to apply this methodology to near-field modeling by including distance as a third dimension during measurements and construction of output mapping functions. For nonlinear behaviors that can be described by tracing domains over which eigenvalues display smooth inflection and branching patterns (such as in mode bifurcation phenomena), time-varying eigenvalues and time-varying output mapping functions could be incorporated. For modeling fingering transitions in the clarinet, a simple interpolation-facilitated scheme for cross-fading between linear models may suffice as presented by Maestre *et al.* (2018) for direction-invariant single-output simulation.

Beyond obvious applications for the interactive use of real-time physical modeling synthesis schemes (Bilbao *et al.*, 2020; Karjalainen, 2008; Maestre *et al.*, 2017a, 2018;

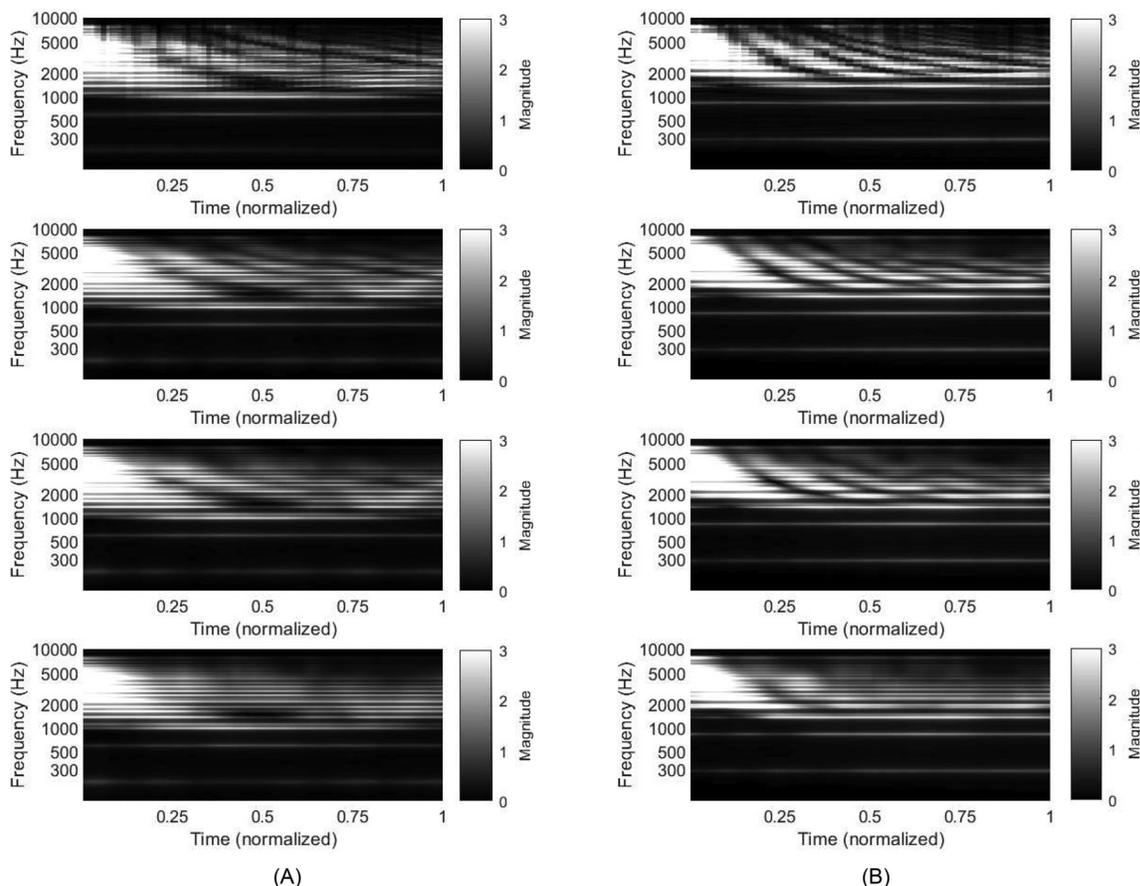


FIG. 10. Time-varying operation of clarinet directivity models of reduced order, simulating the linear trajectory of an ideal microphone on the space of orientation angles over 512 time steps from  $(\theta = 0, \varphi = 0.1\pi)$  to  $(\theta = 2\pi, \varphi = 0.8\pi)$ . (A) G3 fingering via a reference time-invariant model with  $J = 31$  vibration modes ( $M = 62$ ). (B) C4 fingering via a reference time-invariant model with  $J = 33$  vibration modes ( $M = 66$ ). [(A),(B)] (Top) Original frequency response, obtained by nearest neighbor; (second) reduced-order model of order  $M_\xi = 24$  via bilinear interpolation over a table resynthesized by SH decomposition; (third) reduced-order model of order  $M_\xi = 16$  via bilinear interpolation over a table resynthesized by SH decomposition; and (bottom) reduced-order model of order  $M_\xi = 8$  via bilinear interpolation over a table resynthesized by spherical harmonic decomposition.

Rabenstein *et al.*, 2007; Smith, 1992) to render spatialized music, there is real potential for use in general-purpose virtual reality environments and computer animation. Often discarded in interactive operation of geometric acoustic frameworks because of the increased computational cost, the simulation of sound source directivity could find in this method an opportunity for flexible integration. In view of the need for simulating not only direct-field wavefronts but also reflected wavefronts, the possibility of simultaneously running directivity models of different orders could enable the intelligent allocation of computational resources by interactively switching among models, depending on the relevance presented by the individual wavefronts. Moreover, propagating state variables instead of wavefronts (Maestre *et al.*, 2019) may prove computationally advantageous in some situations.

#### ACKNOWLEDGMENTS

This research was supported by the Social Sciences and Humanities Research Council of Canada, Insight Grant No. 435-2018-0720, and the Center for Interdisciplinary Research in Music Media and Technology at McGill

University. Thanks to Harish Venkatesan for assistance during measurements and to the anonymous reviewers for their insightful comments.

Bank, B. (2018). "Multichannel equalization and crosstalk cancellation using fixed-pole IIR filters," *J. Audio Eng. Soc.* **66**(11), 901–909.

Bank, B., and Karjalainen, M. (2010). "Passive admittance matrix modeling for guitar synthesis," in *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*.

Behler, G., Pollow, M., and Vorländer, M. (2012). "Measurements of musical instruments with surrounding spherical arrays," in *Proceedings of Acoustics Conference*.

Bilbao, S., Ahrens, J., and Hamilton, B. (2019). "Incorporating source directivity in wave-based virtual acoustics: Time-domain models and fitting to measured data," *J. Acoust. Soc. Am.* **146**, 2692–2703.

Bilbao, S., Desvages, C., Ducceschi, M., Hamilton, B., Harrison-Harsley, R., Torin, A., and Webb, C. (2020). "Physical modeling, algorithms, and sound synthesis: The NESS project," *Comput. Music J.* **43**(2–3), 15–30.

Bissinger, G., Williams, R. G., and Valdivia, N. (2007). "Violin *f*-hole contribution to far-field radiation via patch near-field acoustical holography," *J. Acoust. Soc. Am.* **121**(6), 3899–3906.

Bodon, K. J. (2016). "Development, evaluation, and validation of a high-resolution directivity measurement system for played musical instruments," M.Sc. thesis, Brigham Young University, Provo, UH.

Canclini, A., Antonacci, F., Tubaro, S., and Sarti, A. (2020). "A methodology for the robust estimation of the radiation pattern of acoustic sources," *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **28**, 211–224.

- Chaigne, A., and Kergomard, J. (2016). *Acoustics of Musical Instruments* (Springer, New York).
- Chaitanya, C. R. A., Raghuvanshi, N., Godin, K. W., Zhang, Z., Nowrouzezahrai, D., and Snyder, J. M. (2020). "Directional sources and listeners in interactive sound propagation using reciprocal wave field coding," *ACM Trans. Graph.* **39**(4), 44.
- Cook, P., and Trueman, D. (1998). "A database of measured musical instrument body radiation impulse responses, and computer applications for exploring and utilizing the measured filter functions," in *Proceedings of the International Symposium on Musical Acoustics*.
- Escolano, J., and López, J. J. (2007). "Directive sources in acoustic discrete-time domain simulations based on directivity diagrams," *J. Acoust. Soc. Am.* **121**, EL256–EL262.
- Fletcher, N. H., and Rossing, T. (1998). *The Physics of Musical Instruments* (Springer, Berlin).
- Grothe, T., and Kob, M. (2013). "Investigation of bassoon directivity," in *Proceedings of the Stockholm Music Acoustics Conference*.
- Grothe, T., and Kob, M. (2019). "High resolution 3D radiation measurements on the bassoon," in *Proceedings of the International Symposium on Musical Acoustics*.
- Hacıhabiboğlu, H., Günel, B., and Kondoz, A. M. (2007). "Source directivity simulation in digital waveguide mesh based room acoustics models," in *Proceedings of the AES 30th International Conference on Intelligent Audio Environments*.
- Härmä, A., Karjalainen, M., Savioja, L., Välimäki, V., Laine, U. K., and Huopaniemi, J. (2000). "Frequency-warped signal processing for audio applications," *J. Audio Eng. Soc.* **48**(11), 1011–1031.
- Huopaniemi, J., Karjalainen, M., Välimäki, V., and Huutilainen, T. (1994). "Virtual instruments in virtual rooms—a real-time binaural room simulation environment for physical models of musical instruments," in *Proceedings of the International Computer Music Conference*.
- James, D., Barbic, J., and Pai, D. K. (2006). "Precomputed acoustic transfer: Output-sensitive, accurate sound generation for geometrically complex vibration sources," *ACM Trans. Graph.* **25**(3), 987–995.
- Karjalainen, M. (2008). "Efficient realization of wave digital components for physical modeling and sound synthesis," *IEEE Trans. Audio, Speech, Lang. Process.* **16**(5), 947–956.
- Lefebvre, A., and Scavone, G. P. (2011). "A comparison of saxophone impedances and their playing behavior," in *Proceedings of the Forum Acusticum Conference*.
- Ljung, L. (1999). *System Identification: Theory for the User*, 2nd ed. (PTR Prentice Hall, Englewood Cliffs, NJ).
- Maestre, E., and Scavone, G. P. (2019). "Creating virtual acoustic replicas of real violins," in *Proceedings of the International Symposium on Musical Acoustics*.
- Maestre, E., Scavone, G. P., and Smith, J. O. (2013). "Digital modeling of bridge driving-point admittances from measurements on violin-family instruments," in *Proceedings of the Stockholm Music Acoustics Conference*.
- Maestre, E., Scavone, G. P., and Smith, J. O. (2016). "Design of recursive digital filters in parallel form by linearly constrained pole optimization," *IEEE Signal Process. Lett.* **23**(11), 1547–1550.
- Maestre, E., Scavone, G. P., and Smith, J. O. (2017a). "Joint modeling of bridge admittance and body radiativity for efficient synthesis of string instrument sound by digital waveguides," *IEEE Trans. Audio, Speech, Lang. Process.* **25**(5), 1128–1139.
- Maestre, E., Scavone, G. P., and Smith, J. O. (2018a). "High-resolution measurement and simulation of sound source directivity," available at <http://www.music.mcgill.ca/caml/doku.php?id=projects:direct>.
- Maestre, E., Scavone, G. P., and Smith, J. O. (2018b). "Joint modeling of impedance and radiation as a recursive parallel filter structure for efficient synthesis of wind instrument sound," in *Proceedings of the 21st International Conference on Digital Audio Effects*.
- Maestre, E., Scavone, G. P., and Smith, J. O. (2019). "Virtual acoustic rendering by state wave synthesis," in *Proceedings of the EAA Spatial Audio Signal Processing Symposium*.
- Maestre, E., Smith, J. O., and Scavone, G. P. (2017b). "Analysis-synthesis of saxophone input impedances via recursive parallel filters," in *Proceedings of the International Symposium on Musical Acoustics*.
- Maestre, E., Spa, C., and Smith, J. (2014). "A bowed string physical model including finite-width thermal friction and hair dynamics," in *Proceedings of the International Computer Music Conference*.
- Mehra, R., Antani, L., Kim, S., and Manocha, D. (2014). "Source and listener directivity for interactive wave-based sound propagation," *IEEE Trans. Visualization Comput. Graph.* **20**(4), 495–503.
- Meyer, J. (1972). *Akustik Und Musikalische Aufführungspraxis (Acoustics and the Performance of Music)* (Edition Bochinsky, Bergkirchen).
- Otondo, F., and Rindel, J. H. (2004). "The influence of the directivity of musical instruments in a room," *Acta Acust.* **90**(6), 1178–1184.
- Pätynen, J., and Lokki, T. (2010). "Directivities of symphony orchestra instruments," *Acta Acust. Acust.* **96**, 138–167.
- Perez-Carrillo, A., Bonada, J., Patynen, J., and Valimäki, V. (2011). "Method for measuring violin sound radiation based on bowed glissandi and its application to sound synthesis," *J. Acoust. Soc. Am.* **130**(2), 1020–1029.
- Pollow, M., Behler, G., and Maseiro, B. (2009). "Measuring directivities of natural sound sources with a spherical microphone array," in *Proceedings of the Ambisonics Symposium*.
- Rabenstein, R., Petrusch, S., Sarti, A., Sanctis, G. D., Erkut, C., and Karjalainen, M. (2007). "Blocked-based physical modeling for digital sound synthesis," *IEEE Signal Process. Mag.* **24**(2), 42–54.
- Rau, M., Abel, J. S., and Smith, J. O. (2018). "Contact sensor processing for acoustic instrument recording using a modal architecture," in *Proceedings of the International Conference of Digital Audio Effects*.
- Shabtai, N. R., Behler, G. K., and Vorländer, M. (2015). "Generation of a reference radiation pattern of string instruments using automatic excitation and acoustic centering," *J. Acoust. Soc. Am.* **138**, EL480–EL486.
- Shabtai, N. R., Behler, G. K., Vorländer, M., and Weinzierl, S. (2017). "Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments," *J. Acoust. Soc. Am.* **141**(2), 1246–1268.
- Smith, J. O. (1992). "Physical modeling using digital waveguides," *Comput. Music J.* **16**(4), 74–91.
- Smith, J. O., and Abel, J. S. (1999). "Bark and ERB bilinear transforms," *IEEE Trans. Speech Audio Process.* **7**(6), 697–708.
- Söderström, T., and Stoica, P. (1989). *System Identification* (Prentice-Hall International, Herdforshire).
- Sparks, M. (1967). "State-space physics," *Phys. Educ.* **2**, 327–333.
- Voronoi, G. (1908). "Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites" ("New applications of continuous parameters in the theory of quadratic forms. First memory. On some properties of perfectly positive quadratic forms."), *J. die Reine Angew. Math.* **133**, 97–178.
- Wang, J.-H., and James, D. L. (2019). "KleinPAT: Optimal mode conflation for time-domain precomputation of acoustic transfer," *ACM Trans. Graph.* **38**(4), 122.
- Weinreich, G. (1982). "Theory and technique of violin radiativity measurements," *J. Acoust. Soc. Am.* **71**, 42–50.
- Weinzierl, S., Vorländer, M., Behler, G., Brinkmann, F., Coler, H. v., Detzner, E., Krämer, J., Lindau, A., Pollow, M., Schulz, F., and Shabtai, N. R. (2017). "A database of anechoic microphone array measurements of musical instruments," available at <http://dx.doi.org/10.14279/depositon-5861.2> (Last viewed September 8, 2000).
- Williams, E. G. (1999). *Fourier Acoustics* (Elsevier, Cambridge, MA).
- Zotter, F. (2009). "Analysis and synthesis of sound radiation with spherical arrays," Ph.D. dissertation, University of Music and Performing Arts, Graz, Austria.