Virtual acoustic rendering by state wave synthesis

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

VIRTUAL ACOUSTIC RENDERING BY STATE WAVE SYNTHESIS

Esteban Maestre
CAML, McGill University
esteban@music.mcgill.ca

Gary P. Scavone
CAML, McGill University
gary@music.mcgill.ca

Julius O. Smith
CCRMA, Stanford University
jos@ccrma.stanford.edu

ABSTRACT

We present State Wave Synthesis (SWS), a framework for the efficient rendering of sound traveling waves as exchanged between multiple directional sound sources and multiple directional sound receivers in time-varying conditions. We introduce a mutable state-space system modal formulation through which input/output matrices change size and coefficients in time-varying conditions, and perceptually motivated designs are possible. SWS enables the accurate simulation of frequency-dependent source directivity and receiver directivity, provides means for simulating frequency-dependent attenuation of propagating waves, and allows an alternative realization through which state variables are treated as propagating waves.

1. INTRODUCTION

Among the challenges of Virtual Reality object-based spatial audio, directionality of sound sources and listeners remains a capstone. In fact, the quest for efficient techniques to model and simulate Head-Related Transfer Functions (HRTF) has arguably been among the most popular in the field. In virtual environments that allow for multiple moving sources, a *classic* form for interactive HRTF simulation requires a database of directional impulse or frequency responses, run-time interpolation of responses, and a time- or frequency-domain convolution engine. Given the predominantly minimum-phase nature of HRTF [1], improved interpolation quality is often achieved if interaural excess-phase is modeled separately as a function of direction. With this form, which is employed in two recently reported systems [2, 3], it is straightforward to swap between personalized HRTF databases [4, 5]; however, one run-time interpolation-convolution channel is required per source wavefront. To reduce the computational needs for each channel, techniques have been studied to simulate HRTF via IIR filters [6], leading to one independent IIR filter per wavefront. Despite the reduction, run-time interpolation of generic IIR filter coefficients is a hard task and leads to artifacts. A number of linear decomposition methods have been used to separate HRTF into parallel basis functions, involving PCA or ICA [7], SVD [8], or Spherical Harmonics (SH) [9, 10]. In turn, the favorable qualities of these parallel models have led to interactive rendering schemes where the parallel basis functions are fixed in number and each of them is implemented by convolution. Following the common-pole observations advanced in [11], state-space models of HRTF have been proposed where size and coefficients of the input matrix are fixed during simulation [12, 13]; though reporting lower computational costs than convolution systems, the size of the input matrix increases with spatial resolution as each system input is statically associated to a fixed direction; also, the methods proposed for joint estimation of transition and input matrices do not allow perceptually-motivated designs as those employed for IIR binaural filters [6]. With respect to efficient simulation of source directivity, some recent works have instead focused on proposing systematic approaches to obtain model-based sound radiation fields using the Equivalent Source Method (ESM) and SH [14]. By low-order SH applied to ESM and pre-computations on fixed virtual scenes [15], frequency responses incorporating source and listener directivity up to around 1 kHz can be generated at rates of 10-15 Hz.

In the context of virtual acoustic simulation relying on traveling wave rendering as dictated by path-tracing methods we present State Wave Synthesis (SWS), a time-domain, convolution-free framework for rendering traveling waves between moving sources and listeners. We introduce a mutable state-space modal formulation that enables simulating directivity of both sources and listeners in terms of input and output matrices of time-varying size and coefficients. We work by first identifying eigenvalues and then constructing time-varying input/output matrix models, thus allowing perceptually-motivated frequency resolutions. The most basic form of SWS comprises three main components: mutable state-space models in modal form used to represent sound source and receiver objects, wave propagators in the form of fractional delay lines and attenuation elements, and input and output mapping functions that facilitate interfacing objects to wave propagators. While the most relevant strength of SWS is its ability to efficiently simulate the directivity of sources and receivers, it also offers means to simulate frequency-dependent attenuation of propagating waves by manipulating the state variables of source object models, and enables a convenient reformulation through which sound propagation is simulated by propagating state variables of source object models.

2. SOURCE AND RECEIVER OBJECTS

Each sound source or receiver object is modeled as a time-varying dynamical system and simulated in terms of a *mutable* state-space model. We use the term *mutable state-space model* to refer to a state-space model for which both its number of input or output variables and their associated input or output vectors mutate dynamically. In such a con-

doi:10.25836/sasp.2019.05

text, we impose that models correspond to strictly proper transfer functions expressible in state-space modal form. Then we write the discrete-time update relation of a mutable state-space model of an object as

$$\begin{aligned} \underline{s}[n+1] &= \underline{\lambda} \odot \underline{s}[n] + \sum_{p=1}^{P} \underline{b}^p[n] x^p[n] \\ y^q[n] &= \underline{c}^q[n]^T \underline{s}[n], \end{aligned} \tag{1}$$

where n is the time index, "$\odot$" denotes element-wise vector multiplication, $\underline{s}[n]$ is a vector of M state variables, $\underline{\lambda}$ is the vector of M system eigenvalues, $x^p[n]$ is the p-th input (a scalar) of those existing at time n, $\underline{b}^p[n]$ is its corresponding length-M vector of input projection coefficients, $y^q[n]$ is a q-th system output (a scalar) obtained as a linear projection of the state variables, and $\underline{c}^q[n]$ is the corresponding length-M vector of output projection coefficients. We refer to the q-th product

$$\underline{c}^q[n]^T \underline{s}[n] \tag{2}$$

as the q-th *output mapping*. The output mapping coefficient vectors enable the projection of the state space of the model onto the output space. We refer to the p-th product

$$\underline{b}^p[n] x^p[n] \tag{3}$$

as the p-th *input mapping*. The input mapping coefficient vectors enable the projection of the input space onto the state space of the model. Note that, as opposed to the classic, fixed-size matrix-based state-space model notation, here we resort to a more convenient vector notation because both the number of inputs or outputs and the coefficients in their corresponding projection vectors (i.e., the numerator coefficients of an equivalent parallel system) are allowed to change (*mutate*) dynamically. In a first basic form, source objects are represented as mutable state-space models for which their outputs are mutable but their inputs are non-mutable (i.e., a fixed number of inputs and input projection coefficients); conversely, receiver objects are represented as mutable state-space models for which their inputs are mutable but their outputs are non-mutable (i.e., a fixed number of outputs and output projection coefficients). However, the framework allows object models for which both inputs and outputs are mutable. Models are constructed by first identifying or defining a set of eigenvalues, and then designing time-varying input/output matrix models via *input/output mapping functions* as outlined below. This two-step procedure allows perceptually-motivated designs.

3. INPUT AND OUTPUT MAPPING FUNCTIONS

Input and output mapping functions enable the mutability of inputs or outputs of a source or receiver object in terms of a number of dynamically changing coordinates associated to those inputs or outputs. For example, the coordinates associated to an input of a sound receiver object may refer to the position or orientation from which the receiver object is excited by a sound wave. As the key elements in mapping functions, *projection models* enable the approximation of the distribution of input and output projection coefficient values over the space of input and output coordinates of an object. Mapping functions employ such projection models to estimate projection coefficients. Without loss of generality, for now we associate input mapping functions to receiver object models and output mapping functions to source object models.

The input mapping function S^+ of a receiver object estimates the input vector $\underline{b}^p[n]$ of projection coefficients corresponding to its p-th input, as a function of an input projection model $\mathcal{B}$ and a vector $\underline{\beta}^p[n]$ of input coordinates. This can be expressed as

$$\underline{b}^p[n] = S^+\big(\mathcal{B}, \underline{\beta}^p[n]\big). \tag{4}$$

Analogously, the output mapping function S^- of a source object estimates the vector $\underline{c}^q[n]$ of output projection coefficients corresponding to the q-th system output, as a function of an output projection model $\mathcal{C}$ and a vector $\underline{\zeta}^q[n]$ of output coordinates. This is expressed as

$$\underline{c}^q[n] = S^-\big(\mathcal{C}, \underline{\zeta}^q[n]\big). \tag{5}$$

Mapping functions may be devised from arbitrary designs or from discrete measurement data. Data-driven construction of projection models enables transforming discrete sets of known projection coefficients (designed, or estimated from measurements) into multivariate continuous functions over the space of the input or output coordinates of an object. This allows having a continuous, smooth time-update of projection coefficients while, for instance, objects change positions or orientations. Notwithstanding the possibility of formulating projection models by way of elaborate modeling methods (e.g., regression in terms of basis functions of different kinds), interpolation of known coefficient vectors may remain cost-effective because only look-up tables are needed. If constrained by memory it should be possible to encode the distribution of projection coefficients via SH instead of storing look-up tables, but this would incur an additional computational cost during regression.

4. WAVE PROPAGATION

A sound wave propagating from the q-th output $y^q[n]$ of a source object to the p-th input $x^p[n]$ of a receiver object may suffer frequency-independent delay induced by the traveled distance, distance-related frequency-independent attenuation induced by wave propagation along the path, and frequency-dependent attenuation due to obstacle interactions (e.g., reflection, transmission, diffraction) or other attenuation causes along the path. A simple model representing these three phenomena can be formulated as

$$x^p[n] = \alpha[n]\, y^q\big[n - l[n]\big] * \chi[n], \tag{6}$$

where $\alpha[n]$ is the scaling factor associated to frequency-independent attenuation, $l[n]$ is the number of delay samples corresponding to the traveled distance given the wave

doi:10.25836/sasp.2019.05

propagation speed and the simulation sampling rate, and $\chi[n]$ is the impulse response of a linear system that models the accumulated frequency-dependent attenuation characteristic $\chi(\omega)$ derived from any obstacle interactions happening along the path or other attenuation causes. Note that $\alpha[n]$, $l[n]$, and $\chi[n]$ may all be time-varying. If the effect of the system characterized by $\chi[n]$ is approximated by a low-order digital filter, a wave propagator responds to a more *classic* structure, for which time-varying frequency-attenuation should be handled by retrieving and/or slewing the coefficients of such low-order filter. Below we will see that, thanks to the state-space formulation, frequency-dependent attenuation can be approximated by scaling the state variables of directional source object models.

5. FREQUENCY-DEPENDENT ATTENUATION VIA STATE ATTENUATION

As an alternative to designing, implementing, and managing digital filters with the sole purpose to model frequency-dependent attenuation due to propagation or obstacle interactions along a path, if the eigenvalues of an object model are conveniently distributed and their associated low-pass (positive real eigenvalue), band-pass (complex-conjugate eigenvalue pair), or high-pass (negative real eigenvalue) components cover representative frequency bands, it is possible to approximate the propagation-induced frequency-dependent attenuation by simply attenuating its state variables at the time of projection. We describe this by using a source object as an example.

For a sound wave traveling from the q-th output of a source object model to the p-th input of a receiver object model, a desired propagation-induced frequency-dependent attenuation characteristic $\chi(\omega)[n]$ may be approximated in terms of a length-M vector $\underline{\gamma}^q[n] = (\gamma_1^q[n], \ldots, \gamma_m^q[n], \ldots, \gamma_M^q[n])$ of source state variable attenuation factors, applied prior to the q-th output projection. In this way, the q-th sound wave $y^q[n]$ departing from the source object model already incorporates the desired attenuation for the path. The new output update relation becomes

$$y^q[n] = \underline{c}^q[n]^T \big(\underline{\gamma}^q[n] \odot \underline{s}[n]\big) \tag{7}$$

where coefficients in $\underline{\gamma}^q[n]$ are real and satisfy $\gamma_m^q[n] \leq 1 \, \forall m = 1, \ldots, M$. With this, the wave propagator relation reduces to

$$x^p[n] = \alpha[n]\, y^q\big[n - l[n]\big]. \tag{8}$$

The system component in charge of estimating the vector $\underline{\gamma}^q[n]$ of state attenuation coefficients is what we refer to as a *state attenuation function*. To estimate the state attenuation coefficients, the state attenuation function may simply sample the desired frequency-dependent attenuation characteristic $\chi(\omega)$ at M frequencies ω_m each corresponding to the natural frequency associated to the m-th eigenvalue of the model. Besides enabling a simplification of the overall implementation provided that a directional source object model presents a convenient set of eigenvalues, this allows the path attenuation function to be easily updated at run-time while the path is still active, e.g., to enable time-varying attenuation properties.

6. STATE WAVE FORM

Through an alternative formulation that we name the *state wave form*, the exact same results can be obtained if eliminating the delay lines of the sound wave propagators and instead treating the source object state variables as propagating waves. To outline this important, yet simple transformation of the system, let us first assume that frequency-dependent attenuation is approximated via low-order digital filters at the end of each propagator delay line. By attending to Equation (1), it should be noted that a traveling wave departing from an object only depends on the state variables of the object model and the vector of coefficients involved in the output projection. Once the output projection is executed, the resulting wave is fed into a fractional delay line for propagation. Let us assume that a sound-emitting object model is feeding a traveling wave $y^q[n]$ into a fractional delay line, and let us refer to the output signal of such delay line as the *delayed traveling wave* signal $d^q[n] = y^q\big[n - l[n]\big]$ with $l[n]$ the fractional sample delay of the line. The delayed traveling wave signal $d^q[n]$ can be expressed in terms of the state variable vector $\underline{s}[n]$ and the output projection coefficient vector $\underline{c}^q[n]$ via

$$d^q[n] = \underline{c}^q\big[n - l[n]\big]^T \underline{s}\big[n - l[n]\big], \tag{9}$$

where $l[n]$ is the delay line length, and $\underline{c}^q\big[n - l[n]\big]$ and $\underline{s}\big[n - l[n]\big]$ are delayed versions of the output coefficient vector and the state variable vector respectively. Since the delayed coefficient vector $\mathbf{c}^q\big[n - l[n]\big]$ can be estimated by the output mapping function given the delayed output coordinates $\underline{\zeta}^q\big[n - l[n]\big]$, i.e.,

$$\underline{c}^q\big[n - l[n]\big] = S^-\Big(\mathcal{C}, \underline{\zeta}^q\big[n - l[n]\big]\Big), \tag{10}$$

the delayed traveling wave $d^q[n]$ can be obtained via

$$d^q[n] = S^-\Big(\mathcal{C}, \underline{\zeta}^q\big[n - l[n]\big]\Big)\underline{s}\big[n - l[n]\big] \tag{11}$$

in terms of *delayed state variables* $\underline{s}\big[n - l[n]\big]$ and *delayed coordinates* $\underline{\zeta}^q\big[n - l[n]\big]$. Thus, instead of propagating traveling waves as emitted by source objects, the system propagates the state variables and the position/orientation coordinates of those objects. If we now assume that frequency-dependent attenuation is approximated via state attenuation, each traveling wave arriving to a sound-receiving object is simply obtained by tapping from the emitting object state variable delay lines and coordinate delay lines at a desired position $l[n]$, and sequentially performing the operations for state attenuation and output projection. This form, which may incur an increase in the cost induced by fractional delay, can be advantageous in diverse application contexts because it eliminates the need for allocating and deallocating delay lines associated to individual propagation paths as traced in dynamically changing scenes. Anecdotally, with this formulation

it could be possible to approximate wave dispersion by tapping at slightly different positions during fractional delay interpolation of the delayed state variables.

7. EXAMPLE MODELS

To provide some simple and illustrative example models, we choose a three-dimensional spatial domain where sound waves propagate as spherical waves radiated from a sound emitting object, propagating in any outward direction from a sphere of finite size representing the object. The direction and position of wave emission by the source are encoded by two angles of three-dimensional spherical coordinates, and constant radius. An equivalent assumption is made for the receiver object: sound waves from a source at a given distance are received from any direction, encoded by two spherical coordinate angles. To portray the presumably more difficult case of modeling real objects as opposed to numerical models of objects, we choose a real acoustic violin as the source object, and a real human body as the receiver object. We demonstrate frequency-dependent attenuation via state attenuation of the acoustic violin model.

7.1 Source object: acoustic violin

An acoustic violin was measured in a low-reflectivity chamber, exciting the bridge with an impact hammer and measuring the sound pressure with a microphone array. The transversal horizontal force exerted on the bass-side edge of the bridge was measured, and defined as the only input of the system. As for the outputs, the resulting sound pressure signals were measured at 4320 positions on a centered spherical sector surrounding the instrument, with radius 0.75 meters from a chosen center coinciding with the middle point between the bridge feet. The spherical sector being modeled covered approximately 95% of the sphere. Each measurement position corresponds to a pair (θ, φ) of angles in the vertical polar convention, conforming the output coordinates on a two-dimensional rectangular grid of $60 \times 72 = 4320$ points. Such a grid represents the uniform sampling of a two-dimensional euclidean space whose dimensions are θ and φ. To design the mutable state-space model of the violin, we first impose minimum-phase and then estimate 58 eigenvalues over a warped frequency axis while jointly accounting for all responses. We then define the input matrix of a corresponding *classic* state-space model as a sole, length-58 vector of ones, and estimate the 4320×58 output matrix by solving a least-squares minimization problem. The solution to this problem equivalently provides estimations for $M = 58$ matrices of size 60×72, with each m-th matrix representing the spherical distribution of output projection coefficient values corresponding to the m-th eigenvalue. From the estimated output matrix, we construct an output mapping function that performs bilinear interpolation of output coefficients over the two-dimensional space of output coordinates, i.e., over the two-dimensional space of angles (θ, φ). To demonstrate the behavior of the source model at run-time, we slew the output coordinates of an outgoing wave as captured by an ideal microphone lying on the sphere surrounding the source object. Assuming ideal excitation of the violin bridge, we simulate a continuous linear motion of the ideal microphone on the sphere, from initial position at ($\theta = 0.69$ rad, $\varphi = 4.71$ rad) to a final position at ($\theta = -1.48$ rad, $\varphi = -0.52$ rad). To illustrate the quality and smoothness of the achieved result, in Figure 1 we compare the measured responses (nearest-neighbor) and the modeled responses as obtained from bilinear interpolation of output projection coefficients.

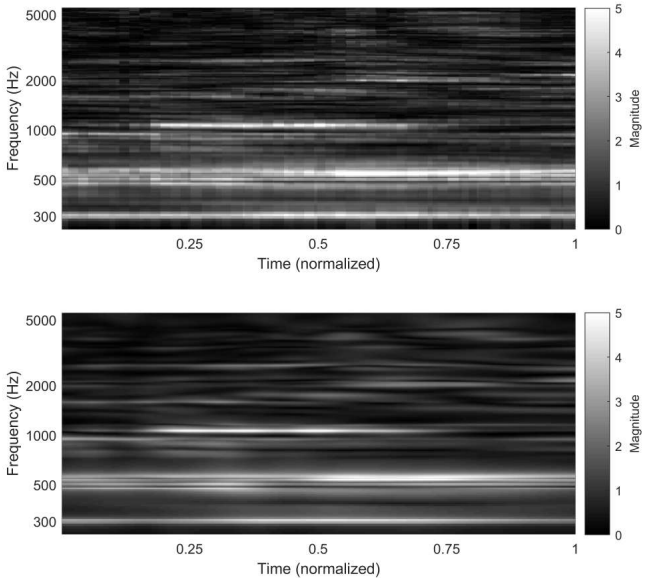

Figure 1. Example modeling (M = 58) of a real acoustic violin as a source object with a spherical sector of 0.75-meter radius as output space expressed in vertical polar coordinates, and the bridge force signal as its only input. Input-output magnitude frequency response as obtained from exciting the model in time-varying conditions: continuous linear motion of an ideal microphone on the sphere, from initial position at ($\theta = -0.69$ rad, $\varphi = -0.34$ rad) to final position at ($\theta = 5.06$ rad, $\varphi = 1.39$ rad). Top graph: nearest-neighbor measurement; bottom graph: model.

7.2 Receiver object: HRTF

To demonstrate the modeling of a receiver object we choose a human body sitting in a chair, as represented by a high-spatial resolution head-related transfer function set of the CPIC dataset [16]. The data used here comprises 1250 responses obtained from measuring the left in-ear microphone signal during excitation by a loudspeaker located at 1250 unevenly distributed positions on a head-centered spherical sector of 1-meter radius. The spherical sector being modeled covers approximately 80% of the sphere. Each of the 1250 excitation positions corresponds to a pair (θ, φ) of azimuth and elevation angles in a two-dimensional space of input coordinates, expressed in the inter-aural polar convention. Following a warped-frequency design procedure analogous to that employed for the violin, we first design a classic state-space model of 1250 inputs, 36 state variables, and one output. From such a model, we first smooth and uniformly upsample the coordinate input space to form $M = 36$ matrices each of $64 \times 64 = 4096$ coefficient values and again choose bilin-

doi:10.25836/sasp.2019.05

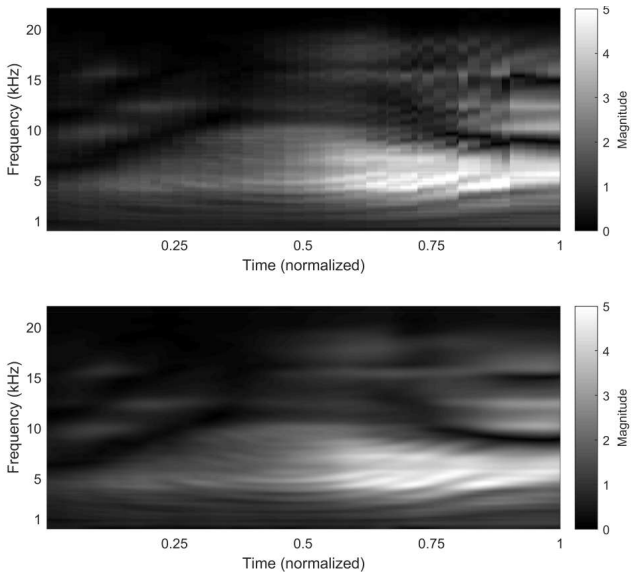

Figure 2. Example modeling (M = 36) of a sitting human body as a receiver object with a head-centered spherical sector of 1-meter radius as input space expressed in inter-aural polar coordinates, and the left in-ear microphone signal as its only output. Input-output magnitude frequency response as obtained from exciting the model in time-varying conditions: continuous linear motion of an ideal source on the sphere, from initial position at ($\theta = 0.69$ rad, $\varphi = 4.71$ rad) to a final position at ($\theta = -1.48$ rad, $\varphi = -0.52$ rad). Top graph: nearest-neighbor measurement; bottom graph: model.

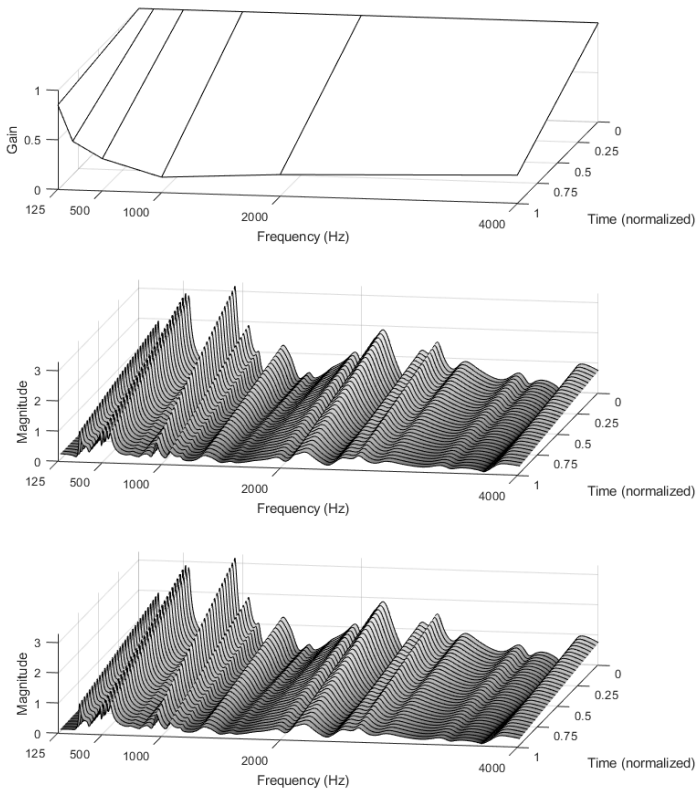

Figure 3. Example modeling of frequency-dependent attenuation via state variable attenuation of a violin model with $M = 58$. For 32 consecutive steps, we modify the frequency-dependent attenuation of an emitted sound wave towards direction ($\theta = -0.34$ rad, $\varphi = 1.39$ rad) by linearly fading from no attenuation to that caused by reflection off a cotton carpet as provided in material tabulated data. Top graph: gain as derived by the attenuation characteristic; middle graph: attenuation simulated via frequency-domain convolution; bottom graph: attenuation simulated via state variable attenuation.

ear interpolation as the input mapping function.

We synthesize the input-output frequency response as obtained from exciting the model in time-varying conditions. For 512 consecutive steps, we modify the input coordinates of an incoming wave as emitted by an ideal source lying on the sphere surrounding the receiver object. We simulate a continuous linear motion of the ideal source on the sphere, from initial position at ($\theta = 0.69$ rad, $\varphi = 4.71$ rad) to a final position at ($\theta = -1.48$ rad, $\varphi = -0.52$ rad). To illustrate the quality and smoothness of the achieved result, in Figure 2 we again compare the measured responses (nearest-neighbor) and the model responses (bilinear interpolation of input coefficients). In the context of binaural rendering, two collocated receiver object models similar to the one demonstrated here could be used, one for each ear. Since the collocated models share position and orientation, the direction coordinates of the incoming traveling wave can be used for both input projection functions. In such context, the required inter-aural time difference can be simulated by tapping from two different positions of the delay line of the incoming wave, and feeding each obtained signal to its respective collocated receiver model.

7.3 Frequency-dependent attenuation

To demonstrate frequency-dependent attenuation via state variable attenuation, we employ the acoustic violin object model described above. Given eight gains $\chi(\omega_k)$, with $\chi(\omega_k) \leq 1 \, \forall k$ and $\omega_1 \cdots \omega_k \cdots \omega_8$ corresponding to octave band frequencies, to estimate the gain required for each of the $M = 58$ state variables of the acoustic violin model (each with natural frequency ω_m), the state attenuation function performs linear interpolation of known gains $\chi(\omega_k)$. We illustrate the time-varying frequency-dependent attenuation by linearly fading, through 32 consecutive steps, between no attenuation (i.e., $\chi_m = 1 \, \forall m = 1, \cdots, M$) and the attenuation caused by a reflection off cotton carpet. This is illustrated in Figure 3, where we compare the results obtained via state attenuation to those obtained by magnitude-only frequency-domain convolution of the departing wave with a response constructed with our scheme.

8. OUTLOOK

We introduced SWS, a time-domain, convolution-free framework that uses a mutable state-space modal formulation for the efficient simulation of both source and receiver directivity in interactive virtual acoustic rendering applications. With our state-space structure, input/output matrices change size and coefficients in time-varying conditions while allowing efficient diagonal forms where, given a state-space order, the computational cost is independent of the spatial resolution. Moreover, our two-stage de-

sign process allows for perceptually-motivated designs via warped-frequency eigenvalue identification as a first step. The framework, which scales well with the number of wavefronts and also enables the simulation of frequency-dependent attenuation, offers a flexible quality-cost trade-off in terms of the order of the state-space models, and allows for a realization where state variables of source objects may be treated as propagating waves.

Straightforward extensions are possible. Though we presented SWS as an early-field renderer, discrete directional components of a simulated diffuse field could be rendered as independent directional wavefronts. The proposed mutable state-space modal representation readily allows for scattering behavior via collocated input and output coordinate spaces and mutability of inputs and outputs, which could be combined with source or receiver behaviors into state-shared hybrid object models. Under linear conditions and relying on the diagonalized modal form of state-space models, it should be possible to dedicate the (mutable) inputs of source objects to serve as a coupling mechanism between a virtual acoustic rendering system and animation-driven rigid-body simulations or physical models. Other attractive routes include simulating effects like near-field and diffraction, or even non-linear source or receiver behaviors in terms of mutable eigenvalues.

Though the provided examples were picked to demonstrate the effectiveness of SWS in accurately simulating highly-directive objects while ensuring smoothness under interactive operation, developments are ongoing and a thorough study of computational cost versus perceptual quality is still required for a fair comparison to other systems. Nevertheless, it remains apparent that simplicity (state-space modal form), efficiency (input and output projections by look-up tables and short vector-scalar multiply-adds), and flexibility (perceptually-motivated frequency designs, state-space order selection, state wave form) make SWS an attractive and resourceful framework for virtual acoustic rendering in diverse application contexts, with some potential for parallel hardware implementations.

9. REFERENCES

[1] J. Nam, M. Kollar, and J. S. Abel, "On the minimum-phase nature of head-related transfer functions," in *AES 125th Convention*, 2008.

[2] D. Poirier-Quinot and B. F. G. Katz, "The anaglyph binaural audio engine," in *AES 144th Convention*, 2018.

[3] M. Cuevas-Rodriguez, L. Picinali, D. Gonzalez-Toledo, C. Garre, E. de la Rubia-Cuestas, L. Molina-Tanco, and A. Reyes-Lecuona, "3d tune-in toolkit: An open-source library for real-time binaural spatialisation," *PLoS ONE*, vol. March, 2019.

[4] C. Hoene, I. C. P. Mejia, and A. Cacerovschi, "Mysofa - design your personal hrtf," in *AES 142nd Convention*, 2017.

[5] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha, "P-hrtf: Efficient personalized hrtf computation for high-fidelity spatial sound," in *IEEE Int. Symposium on Mixed and Augmented Reality*, 2014.

[6] J. Huopaniemi and J. O. Smith, "Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters," in *16th AES Int. Conf. on Spatial Sound Reproduction*, 1999.

[7] V. Larcher, J.-M. Jot, J. Guyard, and O. Warusfel, "Study and comparison of efficient methods for 3d audio spatialization based on linear decomposition of hrtf data," in *AES 108th Convention*, 2000.

[8] J. S. Abel and S. H. Foster, "Method and apparatus for efficient presentation of high-quality three-dimensional audio including ambient effects," *US Patent and Trademark Office*, US5802180, 1998.

[9] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *Journal of the Acoustical Society of America*, vol. 127:2, pp. 823–828, 2009.

[10] M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich, "3d binaural sound reproduction using a virtual ambisonic approach," in *IEEE Int. Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, 2003.

[11] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," *IEEE Trans. on Speech and Audio Processing*, vol. 7:2, pp. 188–196, 1999.

[12] P. Georgiou and C. Kyriakakis, "A multiple input single output model for rendering virtual sound sources in real time," in *IEEE Conf. on Multimedia and Expo*, 2000.

[13] N. H. Adams and G. H. Wakefield, "State-space synthesis of virtual auditory space," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16(5), 2008.

[14] D. L. James, J. Barbic, and D. K. Pai, "Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources," *ACM Trans. on Graphics*, vol. 35:1, pp. 987–995, 2006.

[15] R. Mehra, L. Antani, S. Kim, and D. Manocha, "Source and listener directivity for interactive wave-based sound propagation," *IEEE Trans. on Visualization and Computer Graphics*, vol. 20:4, pp. 495–503, 2014.

[16] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The cipic hrtf database," in *IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, 2001.

doi:10.25836/sasp.2019.05