A hybrid method for creating auralizations of vibroacoustic systems

A. Jared Miller[a)], Scott D. Sommerfeldt[b)], Jonathan D. Blotter[c)] and David C. Copley[d)]

(Received: 24 June 2021; Revised: 5 May 2022; Accepted: 11 August 2022)

A hybrid method for creating a unified broadband acoustic response from separate low-frequency and high-frequency simulation responses is proposed. This hybrid method is ideal for creating simple auralizable approximations of complex acoustic systems. The process consists of four steps: 1) creating separate low-frequency and high-frequency responses of the system of interest, 2) interpolating between the two responses to get a single broadband magnitude response, 3) introducing amplitude modulation to the high-frequency portion of the response, and 4) calculating approximate phase information. Once the appropriate frequency response is obtained, an inverse fast Fourier transform is applied to obtain an impulse response. An experimental setup of an acoustic cavity with one flexible wall is used to validate the hybrid method. The simulated and measured impulse responses are both convolved with various excitation signals, so the validity of the approach could be assessed by listening. Listening tests confirm that the method is able to produce realistic auralizations. The degree of realism is subject to a few limitations, such as pitch differences and dependence on the presence of transients in the excitation signal, but these limitations are incidental and only indirectly related to the proposed method.

Primary subject classification: 76.9; Secondary subject classification: 13.7.1

1 INTRODUCTION

Accurately simulating the vibroacoustic response of a system is valuable in many industries and applications. Various techniques are used to model a system. Lumped parameter models, which simplify the systems into discrete mass, stiffness, and damping elements, are one of the more basic methods. This is often the first type of model one learns in an introductory physics or engineering course on vibration[1]. Despite its relative simplicity, it is frequently used to model vibroacoustic systems with reasonable accuracy. Fahnline and Koopmann[2] use a lumped parameter method to model the acoustic power radiated from a vibrating structure. Karnopp[3] develops lumped parameter models of acoustic filters similar to those found in exhaust systems. Beranek and Mellow[4], as well as Tilmans[5], detail lumped parameter models of various transducers including loudspeakers and microphones.

Lumped parameter estimates are not appropriate for all systems, and there are numerous methods for modeling continuous vibroacoustic systems. Analytical solutions can be derived for some simpler systems, although the extent of this approach is fairly limited[6]. Numerical methods have been developed and are commonly used for vibroacoustic applications where analytical solutions are not feasible. Finite element analysis (FEA) is a method where a larger system is modeled by breaking it into smaller continuous pieces called elements which are connected at nodes and where the interactions occur[7]. This method can be computationally expensive, although advances in computer technology continue to lessen this limitation, and it has become standard due to its accuracy and versatility[8]. While it is most common for structural/mechanical systems, FEA has been used to model many multi-physics-based systems including coupled vibroacoustic systems. Gan et al.[9] use FEA to model sound transmission through the human ear. Nefske et al.[10] examine an FEA formulation for structural-acoustic analysis of the enclosed cavity of an automobile passenger compartment. Everstine[11] provides a review of several FEA formulations used to solve coupled fluid-structure acoustic problems.

The boundary element method (BEM) is a numerical method for solving boundary value problems of partial differential equations. It is, for the most part, more common than FEA for acoustic problems[8]. Kirkup[12] provides a survey of research and applications of the BEM for vibroacoustic

[a)] Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602, USA; email: ajm913@gmail.com.
[b)] Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602, USA; email: scott_sommerfeldt@byu.edu.
[c)] Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602, USA; email: jblotter@byu.edu.
[d)] Caterpillar Inc., Technical Center, Mossville, IL 61552, USA; email: Copley_David_C@cat.com.

problems. The BEM's efficiency for problems where there is a small surface/volume ratio makes it ideal for modeling acoustic radiation from vibrating structures; however, FEA tends to be more appropriate for contained systems[13].

Statistical energy analysis (SEA) is another common method for modeling vibroacoustic systems. In SEA, the complex system is divided into subsystems, and the primary objective is to properly balance the distribution of energy among the subsystems. Within each subsystem, the energy is assumed to be equally distributed among the resonance modes, and the resonances are assumed to be uniformly distributed in frequency within specified frequency bands[14]. Thus, SEA only provides an average level in each subsystem and is more accurate at higher frequencies where a higher modal density occurs[15]. Statistical energy analysis is well established and continues to play a role in ongoing research. Price and Crocker[16] use SEA to model sound transmission between rooms through double panels. Chen et al.[17] propose an affine interval perturbation SEA method to reduce uncertainty in models of a plate-cavity coupled system as well as a simplified launch vehicle fairing.

Each of the above methods has its own advantages/disadvantages, but the overall issue common to all methods is that they are only valid or feasible for limited frequency ranges. For many numerical methods, computation time significantly increases for higher frequency ranges—for example, computation time is proportional to frequency cubed for rectangular acoustic cavities in a modal analysis[18]. Such methods are termed "low-frequency" because they are often not practical for obtaining a response up to higher frequencies (higher frequencies being defined separately for each unique system depending on the geometry and material/fluid properties). In contrast to these low-frequency methods, which become impractical with increasing frequency, energy-based methods such as SEA typically improve with increasing frequency due to higher modal densities.

The research presented here seeks to create a hybrid method, combining a low-frequency method and a high-frequency method to obtain a broadband acoustic response of a vibrating system. The hybrid method is intended to create auralizable responses that can be assessed by listening. The next section provides some additional background about work leading to the hybrid method. Details about the hybrid method follow, along with references to other research attempting to create broadband responses. An experimental setup is used to validate the hybrid method, and results are presented comparing the hybrid method to measurements.

2 DEVELOPMENT

The systems simulated in this paper are meant to represent basic structural/acoustic coupling found in heavy equipment, but the methods developed are generally applicable for many applications. The ultimate objective is to auralize the acoustic response that would be heard by the operator inside the cab of a vehicle or large equipment. As such, the developed method needs to create a broadband response of the vibroacoustic system that can then be auralized and evaluated by listening, unlike many other methods that are only evaluated graphically. The measure of success is then directly related to the perception of the simulated sounds, with the goal of creating sounds that are perceived as "realistic," not necessarily perfect, and avoiding an overall perception of artificialness. The sounds were initially evaluated through listening tests where participants were asked to rate the simulated sounds. A machine learning model was later built to streamline the process, accurately predicting the perceived fidelity or sound quality of the simulated sounds. While briefly mentioned here, the full details of the machine learning model are presented elsewhere[19].

The benefits of accurately simulating the acoustic response of an equipment cab are twofold: first, the acoustic response can be used in a broader simulation that is used to train operators to use the equipment, and second, the model can be used as a design tool to help achieve a desired sound. However, the developed method has implications beyond this specific application and can be used to combine any two traditional methods (one low frequency and one high frequency) to obtain a realistic wide bandwidth acoustic response. It should be noted that the regions of low- and high-frequency are dependent on the system being studied. However, the low-frequency regime can be thought of as the region where the modal density for the system is low, such that the system response can be determined from a limited number of modes. On the other hand, the high-frequency regime corresponds to the region where the modal density is high, such that the response can be considered from a statistical perspective. Often this region is considered to be where there are about 3 or more modes per third octave band.

Simplicity and efficiency are two criteria that guided development of the simulation method. Prior to creating the hybrid method, various methods were tested for creating simple approximations of the acoustic response inside a cab. Two methods emerged as desirable solutions based on these criteria: classical modal analysis (CMA) and SEA. Classical modal analysis was chosen over FEA because it is generally more computationally efficient, which satisfied our criteria of simple and efficient. However, FEA could easily be used if the user was not concerned with achieving increased efficiency.

In CMA, the in vacuo structural modes and rigid boundary acoustic modes are first determined analytically. These independent analytical modes are then combined through spatial coupling coefficients. The final response is then obtained by summing over the total number of modes to be used for the desired frequency bandwidth[20,21]. A convenient matrix formulation of CMA was developed by Kim and Brennan[22] that is based on the impedance/mobility approach. This allows for efficient calculation of coupled responses when the independent analytical modes can be determined. The low-frequency responses presented later in this paper were obtained using this matrix formulation of CMA. If determining the analytical modes is infeasible, FEA could be used to determine the low-frequency responses.

Statistical energy analysis was used to create the high-frequency responses. Statistical energy analysis is very computationally efficient, making it an ideal candidate for the simple approximate model developed in this paper. One major limitation of SEA is that, since it results in average levels, the final responses do not contain the phase information necessary to create the impulse response needed for creating auralization results. The hybrid method outlined below seeks to overcome this limitation and allow for SEA responses to be combined with a low-frequency response to create auralizable broadband acoustic responses.

2.1 Hybrid Method

There are many applications where only considering either low or high frequencies is not enough, and broadband responses are required. This is particularly true when the final response will be auralized, since human hearing spans approximately 20 Hz to 20 kHz, and a reduced frequency range is often perceived as unnatural. Significant research has been done investigating ways to achieve broadband simulations involving acoustic radiation from coupled vibrating structures. No universal method has been found, and it remains an active area of research[23–27]. Many of the proposed methods are quite complex and/or application specific, keeping them from being more widely adopted. Wang et al.[24] use a hybrid approach combining a node-based smoothed finite element method (FEM) and SEA and show good results for several theoretical systems; however, they apply the different methods to separate subcomponents of the system, leaving the SEA portions absent of any phase information. Chronopoulos et al.[25] incorporate a wave FEM with SEA to better account for dispersion in curved shells. Yotov et al.[26] introduce a non-parametric stochastic FEM allowing them to accurately model responses of spacecraft in high-frequency ranges where structures begin to exhibit chaotic behavior and element-based techniques are typically unreliable. Aretz et al.[27] combine FEA, image sources, and stochastic ray tracing to simulate broadband impulse responses. This work is most similar (in objective, not method) to the research presented in this paper, but the method does not achieve the simplicity aimed for here and would be difficult to implement in more complex systems. They provide additional references to similar work, citing limitations and unsatisfactory results in most cases.

There are some established methods that combine low- and high-frequency methods to compute the response of vibroacoustic systems. Certain computer software packages, VA One for example, will simulate complex systems by combining individual components that are each modeled by either FEA or SEA[28]. The user determines which method (FEA or SEA) will be used for each component depending on its geometry and material properties. Such computer simulations can provide accurate responses for complex systems; however, they can become extremely computationally expensive/time-consuming when the response over a large frequency range is desired. Additionally, the energy-based portion of the solution only provides an average level, and it does not capture any resonance or phase information. This becomes problematic if one desires to auralize the simulated response.

There are two main drawbacks in many of the existing broadband solutions. First, the methods are often quite complex. They either require significant computation time or they involve complicated mathematical techniques that are only applicable in specific situations. As previously stated, the goal of this project is to create a simple method to model vibroacoustic systems that is both computationally efficient and simple enough to easily change and apply in various configurations. Of course, there must be a trade-off here: the simpler the model, the less likely it will be able to capture all the complexities of the system. Accordingly, the measure of success is creating a method where the resultant models sound “realistic,” not perfect.

The second drawback is related to the way that many of the existing broadband solutions are evaluated. Plotting the magnitude of the frequency response of the system is the most common way that model accuracy is evaluated. Even when different methods are used in different frequency ranges, the results are often just plotted side by side, without providing any real way to combine the results into a single overall response[29]. This may be sufficient in many instances; however, our main concern is about how the simulated sounds are perceived compared to real sounds. Therefore, our method needs to produce a result that can be auralized. It must be a single response that contains both magnitude and phase information across the frequency range of interest.

The hybrid method developed here seeks to overcome these two problems. In the end, it creates a simple model that produces auralizations that are reasonable approximations of how the real system sounds. There are four steps in the process: 1) creating a separate low-frequency modal response and a high-frequency SEA response of the system, 2) interpolating between the two responses to get a single broadband magnitude response, 3) introducing amplitude modulation to the SEA portion of the response, and 4) calculating approximate phase information. Each of these steps is discussed below and pictured in Fig. 1.

First, two separate responses are calculated, one using a low-frequency method and one using a high-frequency method. For this project, CMA, based on a matrix formulation developed by Kim and Brennan[22], was used to calculate the low-frequency response. Although FEA is probably more commonly used due to its accuracy and ease of implementation with modern software packages, CMA was chosen because of its simplicity and computational efficiency. By way of illustration, a model of an acoustic cavity coupled on one side to a simply-supported vibrating plate (see experimental setup section below) was created with both FEA and CMA. The full finite element forced response mode superposition model took about 44 minutes to run while the classical modal model took only about 1.2 minutes to obtain a frequency response with the same frequency resolution and bandwidth, showing the benefit of using the CMA approach. The high-frequency response was obtained by building an SEA model in the computer modeling software. The SEA response was calculated in one-third octave bands. SEA also meets the simplicity and efficiency criteria.

Second, a single magnitude response was created by interpolating between the separate low- and high-frequency response magnitudes. At this point, only the magnitude response can be obtained because the SEA portion of the response does not contain any phase information. Built-in MATLAB interpolation methods were used to obtain the single unified response. It is important to choose the interpolation method carefully to avoid unexpected results (MATLAB documentation recommends using interp1 with the "pchip" interpolation method when the signal x is not slowly varying)[30]. Determining the crossover frequency, or the point at which to switch from the modal response to the SEA response, is another important consideration in this step. Various crossover frequencies were tried, and it was found that examining the number of modes per frequency band is useful in determining an appropriate crossover frequency. This is discussed in greater depth in the results section, and guidance is provided there on how many modes should be present in a one-third octave band before crossing over to SEA. However, once the crossover frequency was determined, each of the individual responses was truncated; everything above the crossover frequency was discarded from the modal response, and everything below the next one third octave band center frequency was discarded from the SEA response, leaving a gap between the crossover frequency and the next one third octave band center frequency. This gap allowed for a smoother transition between the separate low- and high-frequency responses. The two separate responses were then combined via the interp1 function, which provided for a smooth, continuous curve fit between the two responses. Query points were provided to the interp1 function in 1 Hz intervals, constraining the resolution of the SEA portion to match that of the modal portion. (If a different frequency resolution is used for the low-frequency response, then appropriate query points would be defined to match the low-frequency resolution.)

Third, the SEA response does not capture any information about resonances/antiresonances. This makes for a very smooth unrealistic response. Of course, it is unknown where the resonances/antiresonances would have occurred—a classical modal model or finite element model would be required to know. However, a more realistic response can be obtained by randomly adding amplitude modulation to the SEA response. Although randomly modulating the response will not create peaks at the exact same frequencies as the real system, it was found that it is sufficient to create a more realistic sounding response. This is because the SEA response is only used in a frequency range where the modal density is high.

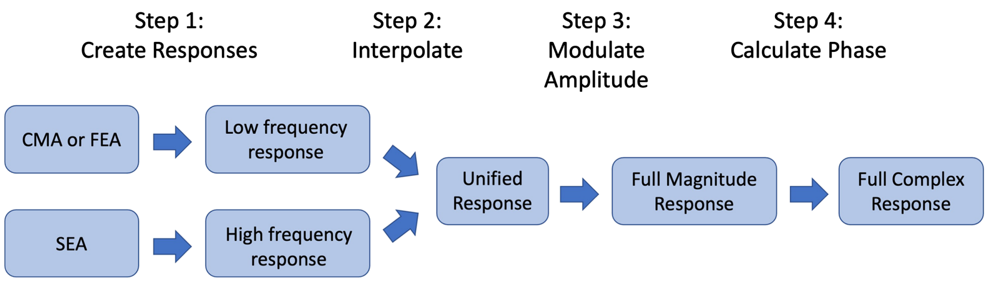

Fig. 1—Diagram representing the steps in the hybrid model process: Step 1) creating a separate low-frequency response and a high-frequency response, Step 2) interpolating between the two responses to get a single broadband magnitude response, Step 3) introducing amplitude modulation to the high-frequency portion of the response, and Step 4) calculating approximate phase information.

In this frequency range, the exact location of the peaks is less important than in the lower frequency range covered by the modal model. The amplitude modulation also has the added benefit of helping create a more realistic phase in the next step. There are two important considerations when creating the amplitude modulation: the magnitude of the modulation and how rapidly the modulation occurs along the frequency axis. The magnitude of the modulation is representative of the damping in the system. Large amplitude modulation represents a system with little damping and results in a high-pitched "metallic" ringing sound in the final simulation. On the other hand, low amplitude modulation represents a system with high damping and results in little to no ringing in the final simulation. As expected, modulating the amplitude of the SEA response only affects the high-frequency ringing (or the ringing in the frequency range where the SEA response is used), while low-frequency ringing is determined by the modal response. Determining the appropriate amplitude to modulate the signal using a computational model can be challenging since it is often difficult to predict the damping in a complex system. However, there are multiple experimental methods one can use to determine the damping in a physical system, which can then be imported into the model. Another method, used here, is to use the magnitude of the peaks and dips in the low-frequency portion of the response to estimate the amplitude by which to modulate the SEA portion, such that a similar variation results in the high-frequency region. This was done by visually inspecting plots of the magnitude of the low-frequency portion, although an algorithmic method could be implemented to streamline the process. In experimenting with this computational method, it was found that it is better to overestimate the damping (underestimate the amplitude of modulation) in the SEA portion of the response, because extensive high-frequency ringing tends to cause the simulated sounds to be perceived as artificial sounding. The amplitude modulation formula used for the results presented in this paper is given by:

$$A' = A * lognrnd(\mu, \sigma) \tag{1}$$

where A' is the modified amplitude, A is the original amplitude, and lognrnd() is a MATLAB function producing lognormal random numbers with parameters μ (mean of logarithmic values) and σ (standard deviation of logarithmic values). The parameter values used to produce the results presented in this paper were $\mu = 0$ and $\sigma = 0.5$. Using $\mu = 0$ corresponds to assuming no D.C. bias in the system response, and $\sigma = 0.5$ resulted in an amplitude modulation that reasonably resembled the modulation in the low-frequency region. Determining how rapidly to modulate the amplitude along the frequency axis is a second concern. Although the exact resonances of the coupled system are not known, the uncoupled natural frequencies of the dominant components can be used to estimate an appropriate density of peaks and dips in the frequency response. In the plate-cavity system described in the experimental setup section below, the resonance frequencies of the plate served as an appropriate approximation.

Fourth, in order to auralize the response, it needs to have phase information as well as magnitude. Therefore, to finalize the SEA based response, an approximate phase needs to be calculated. Significant time was spent experimenting with various ways of creating this approximate phase. Some things that were tried include random modulation similar to the modulation added to the magnitude, calculating a minimum phase via the Hilbert transform, setting the phase at each peak/dip in the magnitude response to π or $-\pi$ respectively and interpolating in-between, and extrapolating from the unwrapped phase of the low-frequency modal response. Although some of the other methods appeared visually better when plotted, the auralized results did not sound natural. Using minimum phase for the phase response was the only method we used that did not create noticeable artifacts when calculating an impulse response and auralizing the results. A minimum phase acoustic system corresponds to one that has the lowest possible time delay for signals being transmitted, and which is also able to be inverted. It is perhaps not surprising that the minimum phase approach worked well, since many natural systems have phase responses that closely resemble minimum phase systems. Thus, while the minimum phase condition may not correspond to the true phase response, it has the desirable property of generally sounding "natural." The process of using the Hilbert transform to create the final response consisted of three parts. First, the magnitude response calculated in the previous step was used to create a two-sided spectrum, since the Hilbert transform expects negative frequencies. Second, the Hilbert transform was used to calculate a minimum phase for the given magnitude response. The formula for calculating the minimum phase is given by

$$\phi(\omega) = -\mathcal{H}[\ln(G(\omega))], \tag{2}$$

where ϕ is the minimum phase, $\mathcal{H}$ represents the Hilbert transform, and G is the two-sided magnitude response. Third, the final complex frequency response was calculated according to

$$\widehat{G}(\omega) = G(\omega) * e^{j\phi(\omega)}, \tag{3}$$

where $\widehat{G}$ is the two-sided complex frequency response, G is the two-sided magnitude response, and ϕ is the minimum phase.

An inverse fast Fourier transform was then applied to the complex frequency response to obtain an impulse response. The impulse response was convolved with various excitation signals to create auralizations, so the validity of the approach could be assessed by listening.

3 EXPERIMENTAL SETUP

A simple coupled structural-acoustic system was built to validate the hybrid method. The system consisted of a rectangular acoustic cavity with five rigid walls and one flexible wall. Similar systems have been studied extensively and used many times to validate new methods[21,31,32]. The rigid walled acoustic cavity was built with a similar method to that used by Kim and Brennan, and the flexible wall was constructed to closely resemble a simply-supported plate, based on a method proposed by Robin et al[22,33].

A diagram of the experimental setup is shown in Fig. 2. Two five-sided boxes were constructed using ½ inch medium-density fiberboard, one larger box and one smaller box designed to sit inside the larger box with a 10-cm gap on all sides. The 10-cm gap between the boxes (including the bottom) was filled with sand so that the inner box acted as a rigid walled acoustic cavity. The inner dimensions of the smaller box were 48 cm × 42 cm × 110 cm. Multiple microphone locations were tested, but for the results shown in this paper, to demonstrate the method, the microphone was located at (20 cm, 18 cm, 63 cm) according to the coordinate system marked in red.

A previously existing aluminum simply supported plate, mounted to a steel frame, was placed on top of the cavity to create the flexible wall (Fig. 3). The plate and cavity were designed to minimize any gaps but prevent touching on the sides, once the plate was placed on top of the cavity. This was done so that the plate dimensions and x-y dimensions of the cavity were equal when modeled while preserving the simply supported nature of the

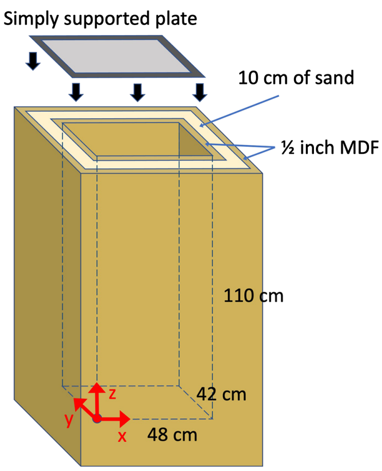

Fig. 2—Diagram of the experimental setup, a simply supported plate coupled to an acoustic cavity.

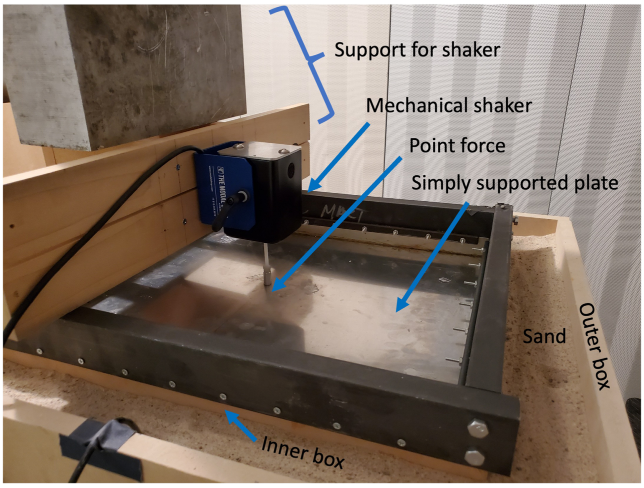

Fig. 3—Photograph of the simply supported plate excited by a mechanical shaker.

plate. The plate was measured to be 3.15-mm thick. The plate was excited by a mechanical shaker at (20 cm, 18 cm, 110 cm), directly above the microphone. A force sensor (not pictured) was attached between the shaker and the plate. The transfer function was measured between the force on the plate and the microphone in the cavity.

4 RESULTS

A model of the plate/cavity experimental setup was built using the hybrid method. The shaker was modeled as a point force, and the microphone was modeled as a point acoustic sensor. For the low-frequency portion of the response, the matrix modal formulation was used[22]. The modal response was calculated up to 2 kHz. Default modeling software material properties for aluminum were used to be consistent with the SEA model: density, 2700 kg/m^3; Poisson's ratio, 0.33; and Young's modulus, 7.1×10^{10} Pa[28]. An airborne sound speed of 340 m/s was used, and the density of air was assumed to be 1.21 kg/m^3, consistent with lab conditions of a room temperature of 20 °C and an elevation of 1400 m. A damping ratio of 0.01 was used, determined by comparing to the measurement since it can be difficult to estimate damping accurately. The high-frequency portion of the response, above 2 kHz, was obtained by creating a SEA model in the modeling software, using all the same parameter values.

The result from the hybrid model is compared to a measurement of the experimental setup in Fig. 4. The pictured response is the transfer function from the input force on the plate to the microphone in the acoustic cavity. These transfer functions were used to calculate impulse responses, which were convolved with various excitation signals (recordings of engine noise and other sounds of interest

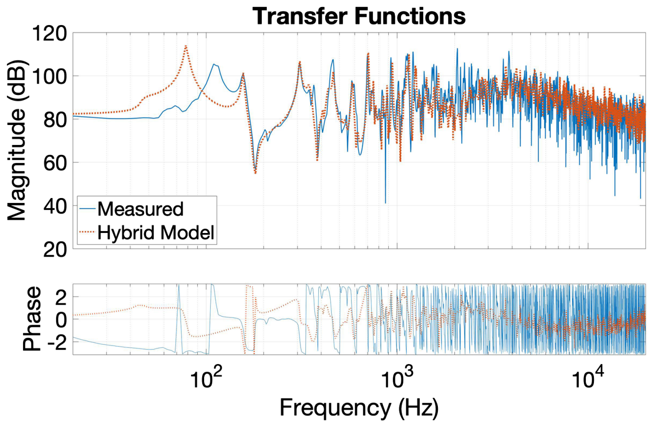

Fig. 4—Full hybrid model result compared to experimental measurement. The transfer functions go from the force on the plate to the microphone in the cavity.

from heavy machinery). Listening to these auralizations is the main way that the validity of the approach was evaluated. However, presenting audio recordings is not possible in a written format, so the frequency responses will be discussed.

One of the most notable features in the frequency response is the mismatch in the frequencies of the lowest peak between the hybrid model and the measurement. The model predicts a peak at 78 Hz, while the measurement showed a peak at 109 Hz. This was concerning and somewhat perplexing considering how well the two matched after the second peak at 156 Hz. Examining the natural frequencies of the plate and cavity individually, one finds that the first peak corresponds exactly to the 1-1 mode of the plate. This is because the 1-1 structural mode is lower in frequency than any of the modes of the acoustic cavity (the first acoustic cavity mode occurs at 155 Hz), so there is no coupling between structural/acoustic modes in this frequency range. The theoretical natural frequency of the 1-1 mode of a simply supported plate with the given material properties is 78 Hz, matching the hybrid model. The discrepancy with the measured response was reconciled by looking at previous measurements of the physical plate separately, not coupled to the acoustic cavity. Those measurements had revealed that the natural frequencies of the experimental plate closely matched those of a theoretical simply supported plate for all the higher modes, but not for the 1-1 mode. The natural frequency of the 1-1 mode was measured to be 109 Hz, exactly matching the measured resonance in the coupled system. Therefore, the discrepancy did not come from an error in the model, but from the inadequacy of the experimental setup in replicating simply supported boundary conditions at these lower frequencies. It is not surprising that the physical boundary conditions do not match the theoretical ones exactly, and it is worth noting that the developers of the method used to construct the simply supported plate also found the largest percent error with the 1-1 mode[33]. Identifying these results as the source of the discrepancy alleviated concerns, and attempting to fix the plate was deemed unnecessary. Applying a high-pass filter with a 100-Hz cutoff frequency to the auralizations proved sufficient in minimizing the differences caused by these mismatching fundamental frequencies.

As previously stated, a full complex frequency response, including both magnitude and phase information, is necessary to transform to the time domain to obtain an impulse response for auralization. Although both are necessary, the magnitude portion of the responses tends to dominate human perception of sound, while the phase plays a secondary role, particularly at higher frequencies[34]. For example, imagine a musical note played from a pair of loudspeakers. Shift the frequency or change the amplitude and people are bound to notice, but shift the phase and there is likely to be no perceptible difference (except in specific circumstances where significant interference occurs). This means that matching the magnitude portion as closely as possible is vital, but finding an appropriate approximation of the phase can be sufficient. By no means does this imply that the phase one uses is completely arbitrary. Out of the infinite number of possible phases, only a small subset will approximate reality close enough to sound natural. It was previously discussed that many ways of constructing an approximate phase were tested, and nearly all of them introduced undesirable artifacts into the final auralizations. Adopting a minimum phase was the one method that preserved the naturalness of the auralizations. Most physical systems are not truly minimum phase although many approximate minimum phase at low frequencies, particularly those systems that are largely characterized by propagating waves. Using minimum phase has nice properties such as preserving causality and invertibility that allow it to produce auralizations without introducing such artifacts. The minimum phase calculated for the hybrid model shown in Fig. 4 is sufficient to create a natural sounding auralization for the system of interest, confirmed via listening tests. To further test the viability of the minimum phase, a new response was created by combining the magnitude of the measured response and the phase of the hybrid model. Auralizations created with this new response were not perceptibly different than those created from the full measured response. This shows that the minimum phase is indeed a satisfactory approach to approximate the real system. It is possible that this may not be the case for all systems, and further investigation would be appropriate to examine the generalizability of using the minimum phase, as well as investigating other methods for calculating an approximate phase.

Many of the simulated auralizations sounded similar to the measurements, although the exact level of similarity was somewhat difficult to assess. Listening tests were conducted to evaluate the similarity, focusing on realism/artificialness. Participants were presented pairs of sounds, a reference measured sound and a simulated sound, and asked to rate whether the simulation sounded artificial when compared to the reference. Eleven listeners participated in the listening tests to capture a variety of perceptions and opinions (the small sample size is recognized as a limitation and further work should be done to determine if these results represent human perceptions as a whole). Further details about the listening tests can be found in a previously published paper[19]. Two trends became apparent when examining listening test responses. First, the perceived pitch of the sounds was dominated by the peaks with the highest magnitude in the frequency response. This had a significant impact on how similar the simulations were perceived compared to the measurements because pitch is one of the main perceptual traits that people tend to focus on when comparing two sounds. However, even though the pitch differences significantly affected perception of the overall similarity of the sounds, they did not significantly affect the perceived realism of the sounds. For example, the peak at 2069 Hz in the measured response shown in Fig. 4 is notably missing from the response of the hybrid model because a 2000-Hz crossover frequency was used for the model. This caused a significant difference in the pitch of the measured sounds versus the sounds created from the hybrid model, resulting in lower ratings for overall similarity. Increasing the crossover frequency to 2100 Hz allowed the model to capture the 2069 Hz peak, resulting in noticeably more similar pitches and therefore better ratings of overall similarity. Despite better ratings for overall similarity between the measurement and the model, there was no difference in the perceived realism of the simulated sounds for the 2100-Hz crossover compared to the 2000-Hz crossover. This shows that exactly matching the dominant peaks is not necessary to create realistic simulations, even if pitch differences are introduced.

Second, the excitation signal had a significant impact on whether the sounds were perceived as realistic or artificial. In particular, it was found that sounds created with input signals containing transients were much more likely to be perceived as realistic, while sounds created from entirely steady state input signals were more likely to be perceived as artificial. This was the case across the board, for both measured sounds and simulated sounds; even measured sounds were more likely to be perceived as artificial if the excitation signal contained no transients.

The crossover frequency is one of the main considerations in the proposed hybrid method. Low-frequency methods usually provide a more accurate result, so theoretically, the crossover frequency should be as high as possible for the best response. However, as mentioned before, low-frequency methods take considerably more computation time. If a simple and efficient model is the goal, the question that naturally arises is, "How low of a crossover frequency is acceptable?" In order to address this question, models were created with crossover frequencies varying from 16 Hz to 2000 Hz at the one third octave band center frequencies. These models were used to create auralizations which were listened to and rated according to perceived realism/artificialness. Although there was some variation, and the perceived realism of the sounds depended on the excitation signal as discussed above, it was found that the lowest "good" crossover frequency was about 630 Hz, as shown in Fig. 5. (Good crossover frequencies were defined based on the ability to produce auralizations that on average sounded more natural than artificial). This does not mean that 630 Hz was necessarily the best or optimal crossover frequency; models with higher crossover frequencies were rated better. Rather, it provides a lower bound for the crossover frequencies that can be used and still retain a sense of realism in the auralizations. The 630 Hz limit is only for this particular setup and would certainly change for other configurations, but it is reasonable to assume that the lower limit would be related in some way to a modal density. The plot in Fig. 5 shows that crossing over at a frequency where each component has 3+ modes per one third octave band, in all parts of the system, preserves the realism of the final auralizations. This seems reasonable since SEA gives a

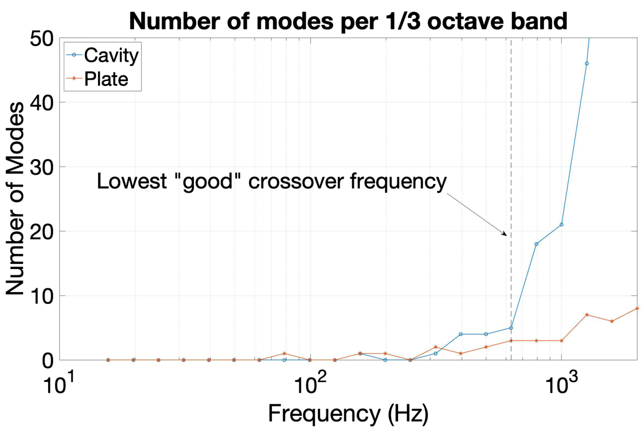

Fig. 5—Number of theoretical modes per 1/3 octave band for the simply supported plate and the acoustic cavity. The lowest "good" crossover frequency for the hybrid model is shown to be approximately 630 Hz.

more accurate result when there are multiple modes per frequency band. Therefore, when computation time is a concern, the number of modes per frequency band can be examined for each component of the system and used to determine an appropriate crossover frequency.

5 CONCLUSIONS

The proposed hybrid method successfully merged a low-frequency response and a high-frequency response into a single response that could be auralized. This allowed for a simple and efficient approximation of the desired acoustic response over a broad frequency range. The auralizations were able to retain a sense of realism, skirting some of the unnatural artifacts prevalent in audio simulation. There were some limiting factors in the level of realism achieved. First, matching the largest peaks in the frequency responses is necessary to create the same pitch, and it was found that pitch differences are a major factor when listeners compare the similarity of two sounds (simulation vs. measurement). Second, the presence/absence of transients in the excitation signal significantly affected the perceived realism of the final auralization. While important considerations, neither of these two limiting factors are directly related to the ability of the hybrid method to produce realistic auralizations. Pitch differences only significantly affect perceived similarity and not necessarily perceived realism, and the presence of transients in the excitation signal is completely situational and unrelated to the developed method. Early on in development, artifacts introduced by the interpolation method or when calculating the approximate phase tended to dominate the perception of the sounds. The final method appears to have overcome these challenges, and they are no longer limiting factors in the achievable realism of the final result.

6 ACKNOWLEDGMENTS

This work was supported by Caterpillar Inc.

7 REFERENCES AND NOTES

1. F. Fahy and D. Thompson, *Fundamentals of Sound and Vibration*, CRC Press, Boca Raton, FL, (2015).
2. J.B. Fahnline and G.H. Koopmann, "A lumped parameter model for the acoustic power output from a vibrating structure", *J. Acoust. Soc. Am.*, **100**(6), 3539–3547, (1996).
3. D. Karnopp, "Lumped parameter models of acoustic filters using normal modes and bond graphs", *J. Sound Vib.*, **42**(4), 437–446, (1975).
4. L.L. Beranek and T.J. Mellow, *Acoustics: Sound Fields and Transducers*, Academic Press, Waltham, MA, (2012).
5. H.A.C. Tilmans, "Equivalent circuit representation of electromechanical transducers: I. Lumped-parameter systems", *J. Micromech. Microeng.*, **6**(1), 157–176, (1996).
6. A.D. Pierce, *Acoustics: An Introduction to Its Principles and Applications*, Springer International Publishing, Cham, Switzerland, (2019).
7. S.S. Rao, *Mechanical Vibrations*, Pearson Education, Inc., Hoboken, NJ, (2017).
8. N.S. Gokhale, S.S. Deshpande, S.V. Bedekar, and A.N. Thite, *Practical Finite Element Analysis*, Finite to Infinite, Maharashtra, India, (2008).
9. R.Z. Gan, B. Feng, and Q. Sun, "Three-dimensional finite element modeling of human ear for sound transmission", *Ann. Biomed. Eng.*, **32**, 847–859, (2004).
10. D.J. Nefske, J.A. Wolf, and L.J. Howell, "Structural-acoustic finite element analysis of the automobile passenger compartment: a review of current practice", *J. Sound Vib.*, **80**(2), 247–266, (1982).
11. G. Everstine, "Finite element formulatons of structural acoustics problems", *Comput. Struct.*, **65**(3), 307–321, (1997).
12. S. Kirkup, "The boundary element method in acoustics: a survey", *Appl. Sci.*, **9**(8), (2019).
13. J.T. Katsikadelis, *Boundary Elements Theory and Applications*, Elsevier, Amsterdam, Netherlands, (2002).
14. R.H. Lyon and R.G. DeJong, *Theory and Application of Statistical Energy Analysis*, Butterworth-Heinemann, Newton, MA, (1995).
15. C.B. Burroughs, R.W. Fischer, and F.R. Kern, "An introduction to statistical energy analysis", *J. Acoust. Soc. Am.*, **101**(4), 1779–1789, (1997).
16. A.J. Price and M.J. Crocker, "Sound transmission through double panels using statistical energy analysis", *J. Acoust. Soc. Am.*, **47**(3), 683–693, (1970).
17. Q. Chen, Q. Fei, S. Wu, and Y. Li, "Statistical energy analysis for the vibro-acoustic system with interval parameters", *J. Aircr.*, **56**(5), 1869–1879, (2019).
18. R.D. Belvins, "Modal density of rectangular volumes, areas, and lines", *J. Acoust. Soc. Am.*, **119**(2), 788–791, (2006).
19. A.J. Miller, S.D. Sommerfeldt, and J.D. Blotter, "Using machine learning to evaluate the fidelity of heavy equipment acoustic simulations", *Appl. Acoust.*, **187**, 108513, (2022).
20. F. Fahy and P. Gardonio, *Sound and Structural Vibration: Radiation, Transmission, and Response*, Elsevier, Oxford, England, (2007).
21. J.T. Du, W.L. Li, H.A. Xu, and Z.G. Liu, "Vibro-acoustic analysis of a rectangular cavity bounded by a flexible panel with elastically restrained edges", *J. Acoust. Soc. Am.*, **131**(4), 2799–2810, (2012).
22. S.M. Kim and M.J. Brennan, "A compact matrix formulation using the impedance and mobility approach for the analysis of structural-acoustic systems", *J. Sound Vib.*, **223**(1), 97–113, (1999).
23. M. Vorländer, "Simulation and auralization of structure-borne sound", in *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Chapter 14 (Springer-Verlag, Berlin, Germany, (2008).
24. G. Wang, Y.X. Zhang, Z.B. Guo, and Z.G. Zhou, "A novel hybrid deterministic-statistical approach for the mid-frequency vibro-acoustic problems", *App. Math. Model.*, **83**, 202–219, (2020).
25. D. Chronopoulos, B. Troclet, M. Ichchou, and J.P. Lainé, "A unified approach for the broadband vibroacoustic response of composite shells", *Compos. B. Eng.*, **43**(4), 1837–1846, (2012).
26. V. Yotov, M. Remedia, G. Aglietti, and G. Richardson, "Non-parametric stochastic FEM / hierarchical matrix BEM for efficient mid-frequency vibroacoustic response estimation," in *25th International Congress on Sound and Vibration*, Hiroshima, Japan, (2018).
27. M. Aretz, R. Nöthen, M. Vorländer, and D. Schröder, "Combined broadband impulse responses using FEM and hybrid ray-based methods", in EAA Symposium on Auralization, Espoo, Finland, (2009).
28. "VA ONE", ESI Group, 2020. [Online]. Available: at: https://www.esi-group.com/products/vibro-acoustics, (accessed December 2020).

29. R. Jiao and J. Zhang, "Vibro-acoustic modeling of a rectangular enclosure with a flexible panel in broad range of frequencies and experimental investigations", *J. Vibroeng.*, **18**(4), 2683–2692, (2016).
30. MATLAB Help Center Documentation. [Online]. Available at: https://www.mathworks.com/help/matlab/interpolation.html?s_tid=CRUX_lftnav, (accessed January 2021).
31. J. Pan and D.A. Bies, "The effect of fluid–structural coupling on sound waves in an enclosure—theoretical part", *J. Acoust. Soc. Am.*, **87**(2), 691–707, (1990).
32. H. Zhang, D. Shi, S. Zha, and Q. Wang, "Vibro-acoustic analysis of the thin laminated rectangular plate-cavity coupling system", *Compos. Struct.*, **189**, 570–585, (2018).
33. O. Robin, J.-D. Chazot, R. Boulandet, M. Michau, A. Berry, and N. Atalla, "A plane and thin panel with representative simply supported boundary conditions for laboratory vibroacoustic tests", *Acta Acust. United Acust.*, **102**, 170–182, (2016).
34. M.-V. Laitinen, S. Disch, and V. Pulkki, "Sensitivity of human hearing to changes in phase spectrum", *J. Audio Eng. Soc.*, **61**, 860–877, (2013).