D. MENEXOPOULOS, P. D. Pestana and J. D. Reiss,
"The State of the Art in Procedural Audio"
J. Audio Eng. Soc., vol. 71, no. 12, pp. 826–848, (2023 December).
DOI: https://doi.org/10.17743/jaes.2022.0108

The State of the Art in Procedural Audio

DIMITRIS MENEXOPOULOS,[1] **PEDRO D. PESTANA,**[2] *AES Member* **AND JOSHUA D. REISS,**[1] *AES Fellow*
(d.menexopoulos@qmul.ac.uk) (pedro.duarte.pestana@gmail.com) (joshua.reiss@qmul.ac.uk)

[1] *Centre for Digital Music, Queen Mary University of London, London, UK*
[2] *Science and Technology Department, The Open University of Portugal (UAb), Lisbon, Portugal*

Procedural audio may be defined as real-time sound generation according to programmatic rules and live input. It is often considered a subset of sound synthesis and is especially applicable to nonlinear media, such as video games, virtual reality experiences and interactive audiovisual installations. However, there is resistance to widespread adoption of procedural audio because there is little awareness of the state of the art, including the diversity of sounds that may be generated, the controllability of procedural audio models, and the quality of the sounds that it produces. The authors address all of these aspects in this review paper, while attempting a large-scale categorization of sounds that have been approached through procedural audio techniques. The role of recent advancements in neural audio synthesis, its current implementations, and potential future applications in the field are also discussed. Review materials are available*.

0 INTRODUCTION

Procedural audio refers to the use of algorithms to dynamically generate audio content in real time, while adapting to changing inputs. It has become an increasingly important aspect of creative sound design, because it enables the creation of immersive and dynamic soundscapes. In recent years, the field of procedural audio has seen significant advancements, with the development of new technologies and techniques aimed at enhancing the audio experience in simulated environments.

The use of procedural audio can be traced back to the early days of video games, in which simple sound effects were generated using basic algorithms. However, the limitations of hardware at the time prevented the widespread adoption of procedural audio. With the advent of more powerful computing systems and the development of new software tools, procedural audio has regained attention as a creative component of game audio design [1].

Procedural audio has also found applications in other areas such as virtual and augmented reality [2–4], audiovisual art installations [5], the automotive industry [6], and even physical therapy [7]. This highlights the versatility of procedural audio, as well as its potential importance for a range of industries and applications.

One of the main advantages of procedural audio in interactive sound design is the ability to adapt in real time to changing inputs, thus creating a unique and dynamic audio experience for each user. Detailed and seamless sonic interactions afforded by procedural audio can contribute toward a heightened sense of immersion within a virtual environment, resulting in a potentially more engaging and memorable experience. Additionally, procedural audio can help toward reducing the file size of audio assets, because the audio content is generated on-the-fly, rather than being pre-recorded and stored in the game or virtual reality engine.

This paper attempts to give an overview of the state of the art in procedural audio, including the diversity of sounds that may be generated, the controllability of procedural audio models and the quality of the sounds that it produces. The authors also provide a large scale categorization of sounds that have been approached through procedural audio techniques and the specific methods used for generating each sound. Potential future applications in the field are also discussed, with an emphasis on the role of recent advancements in neural audio synthesis.

SEC. 1 discusses the background to the field, including previous reviews of procedural audio and related fields. SEC. 2 explains the methodology for finding and presenting the state of the art in procedural audio. The authors' approach aimed for a systematic review of the relevant literature. SEC. 3 attempts a concise taxonomy of sounds produced with procedural techniques, and SEC. 4 attempts to classify the synthesis techniques used to produce those sounds. In SEC. 5, Table 1 documents in detail the connection between the sound classes and synthesis techniques employed to create them, and patterns

---

*https://dmenex.github.io/proceduralaudioreview/

emerging from this relation are pointed out. SEC. 6 describes the design processes that informed the synthesis techniques and parameters used for generating each procedural audio model. The evaluation of procedural audio techniques is discussed in SEC. 7. Emerging and promising procedural audio research, especially related to machine learning, is discussed in SEC. 8. Finally, overall findings are discussed and conclusions are drawn in SEC. 9.

1 BACKGROUND

There have been a few past attempts to review the state of the art in related areas of sound synthesis, thus providing valuable insights into the context within which procedural audio sits and highlighting its potential for future development. Researchers with extensive original contributions to the field of procedural audio include Perry Cook, Andy Farnell, and Davide Rocchesso. Cook is one of the originators of the field and focused primarily on physics-based sound modeling techniques. His book [8] includes many synthesis technique explanations and provides detailed examples of sounds produced by rigid bodies, like strings, bars, and tubes. Rocchesso et al. [9] provide a detailed exploration of the physical principles of sound synthesis and a thorough examination of the various methods for generating and controlling the sound of objects. They also explore the artistic and practical applications of this approach. Farnell coined the term *procedural audio*, and his book *Designing Sound* [10] provides many practical examples covering procedural audio for common sound families approached with various sound synthesis methods.

Strobl et al. [11] produced an early review of techniques for modeling sound textures, in which sound textures may be defined as those sounds whose statistical properties remain fairly constant over a sustained period of time, such as crackling fire, running water, and applause. This work also included a breakdown of synthesis methods and provided some practical examples. Another review of sound texture synthesis techniques was given in [12], with an emphasis on granular approaches such as concatenative synthesis. Notably, this review paper highlighted a lack of evaluation in the field. A more recent review was provided in [13], using the alternative term *audio textures*, giving a history of the field, and discussing analysis methods and applications, in addition to synthesis.

A general review of sound synthesis and analysis techniques was provided in [14]. This review covered speech and music synthesis, as well as sample-based approaches, and hence had a very wide remit and was intended as an introduction to the field. A more recent overview is given in [15], in which the authors reviewed techniques to synthesize sound effects, as used in creative media. Both reviews were aimed at sound designers and sound design researchers, and emphasized definitions, classifications, techniques, and examples.

Several papers have reviewed techniques for generating different aeroacoustic sounds. In [16], Rizzi and Sahai documented the state of the art in auralization, the process analogous to visualization in describing rendering audible (imaginary) sound fields [17], of air vehicle noise. They classified the existing approaches into two main categories: time domain approaches that perform sound synthesis followed by propagation, and frequency domain approaches that perform propagation followed by sound synthesis. In a minor review, Selfridge produced and summarized various procedural audio models for producing aeroacoustic sounds, such as aeolian harp, propeller, and cavity tones, using physically derived equations [18]. Böttcher et al. produced, tested, and reviewed a niche of techniques for creating aerodynamic sounds using the visual programming language Max/MSP and a Wii controller, providing results from a subjective evaluation [19].

In addition to sound textures and aeroacoustics, the generation of soft body sounds (such as crumpling, rubbing of cloth, paper, jelly, or rope) is of particular interest and contains notable challenges due to the complex, non-tonal nature of such sounds. In a focused, yet detailed review, Su et al. examined procedural audio techniques used specifically for generating soft body sounds [20].

Hawley et al. [21] recently reviewed, compared, and contrasted physical modeling and machine learning approaches to musical instrument synthesis. Natsiou et al. [22] focused just on reviewing deep learning approaches, but in the wider context of synthesizing any sound. However, neither paper focuses on methods that adapt to relevant changing inputs, and hence, the approaches they review generally fall outside the context of this paper.

2 RESEARCH METHODOLOGY

2.1 Definition and Scope

The most widely used definition of procedural audio first appeared in [23] and is given as “nonlinear, often synthetic sound, created in real time according to a set of programmatic rules and live input.” The same definition is used herein, but with some additional focus that clarifies or limits the scope.

First, music and speech synthesis approaches are excluded from the review. This is mostly because of the fact that these two areas are already very large and mature. Also, prior research has identified the need for procedural audio models of non-speech, non-music sounds [10, 15, 24].

The authors are also interested in the pure form of procedural audio, in which audio content is generated “from scratch” and does not employ any stored audio recordings. As such, granular synthesis [25], concatenative synthesis [26], and related approaches are excluded, because they require audio samples as inputs, which are then extensively modified to generate new sounds.

The authors note that restricting the review only to techniques with no stored audio is arguable. Fine-grained, sample-based approaches such as granular synthesis allow for control in a statistical fashion, and concatenative synthesis has been shown to be highly effective in sound texture synthesis [12]. However, some authors are very clear on this restriction, noting that “any recording-based method, such as granular synthesis, is not considered as procedural audio" [27]. This restriction also provides a clear dividing line between procedural and sample-based techniques, which is an often-used classification, e.g., [3, 28, 29].

A useful distinction, which both maintains the purest definition of procedural audio and allows for the role of granular, sample-based approaches is given in [30], which stated, "Procedural audio is sound entirely generated using algorithms and synthesis techniques, as opposed to procedural sound design, which uses pre-recorded samples."

Another important consideration is the ability to control meaningful sound parameters in real time. This is a key characteristic of procedural audio, and hence, the authors attempted to only include models that strongly demonstrate this feature. A sound model that simply generates footstep-like sounds without adaptive control, for instance, would not be considered procedural. Nor would it be procedural if one can just adjust abstract parameters of a model, such as coefficients in a Fourier series. Instead, one needs to be able to adjust meaningful parameters such as walking speed, type of shoe, and type of surface while the footsteps are being generated. Thus, current neural synthesis techniques are also not considered procedural because they generally are not adaptive and do not provide meaningful real-time controls. However, they show a lot of promise, and it is likely that procedural neural approaches will emerge in the near future. Thus, they are discussed in SEC. 8.

Sound models that have relevant controllable parameters, but either are non-causal and hence cannot be implemented in real time or are too computationally heavy to run in real time, are excluded.

2.2 Identification and Selection of the State of the Art

The term "Procedural Audio" captures the type of implementation the present authors are going for, but many authors do not use the term. Definitions as broad as "synthesis technique" have been used in the literature, but these would make any search methodology unwieldy and would return an excessive number of results that are not applicable. It was thus decided to use the search strings "Procedural Audio" and the rarer term "Computer Generated Audio" and expand from there. The authors' actions were as follows:

1. Use the search strings in Scopus, Web of Science, and Google Scholar in title, keywords, and abstract. This generated lists of 312, 217, and 544 references, respectively.
2. Collate the titles and eliminate duplicates through a script, for a total of 712 references.
3. Do a first screening round on titles alone, discarding all instances that are clearly unrelated, or distinctly imply non–real-time or sample-based approaches. In this stage, 167 references remained.
4. Find and collect all remaining articles.
5. Do a second screening round on abstracts, again discarding on the basis of the aforementioned criteria. In this stage, 92 articles were selected.
6. Check all articles for type of sounds modeled, synthesis technique, design technique, and evaluation. Discard those that appeared to fit the criteria through previous screening methods but actually did not, based on a complete read. Discard duplicates that were not found by the script used in item 2 above (three articles were discarded in this phase).
7. With the remaining articles, run a forward (cited by) and backward (bibliography) search in Google Scholar through a script. Collate the results and eliminate duplicates. Eliminate all results that had previously been screened. This brought a total of 102 titles.
8. Do one last round of title screening followed by abstract screening followed by checking all articles.
9. During the course of writing, add other articles that have become available or that the authors are made aware of. For reproducibility sake, articles that were included according to this principle are [31–33]. The final number of included articles is 99.

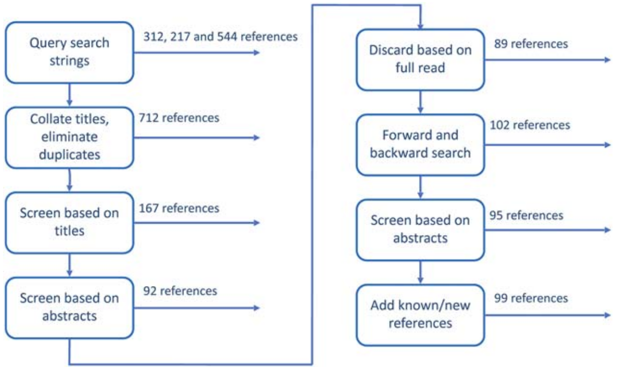

Fig. 1. Flowchart highlighting the study selection process and the number of articles selected at each step.

Fig. 1 provides a flowchart of the study selection process, outlining the steps given above. Throughout all exclusion phases, exclusion criteria were

- Articles that fall outside the scope of the procedural audio definition above (non–real-time, non-controllable, using samples, targeting speech or music),
- Articles that mention a model but do not implement or give an outline of its algorithm, and
- Articles that refer to a previously existing model.

A brief bibliographical analysis on the complete table was performed in order to understand whether this is a field that is widespread enough and growing in time, and to provide an overview of where research is happening. There are 174 authors represented in the works under analysis, 138 of which collaborated on a single article. Only one author contributed to more than ten articles. In terms of where the research is published, the Audio Engineering Society (19 articles in either the journal, conventions, or

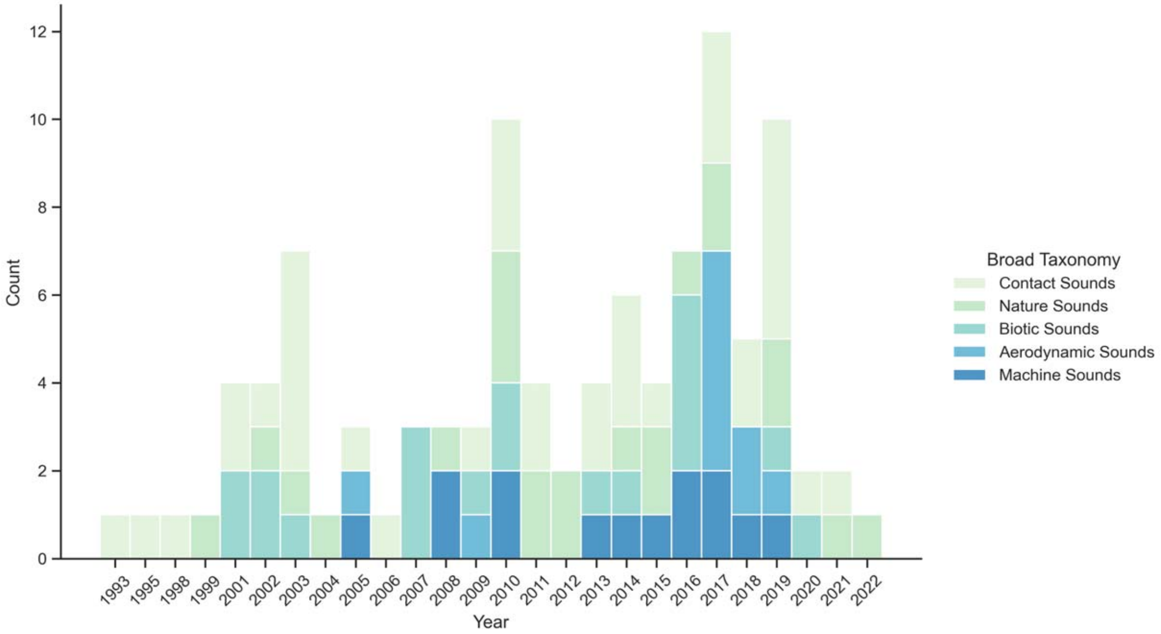

Fig. 2. Number of articles published since 1993 including a practical algorithm in procedural audio, broken down into the main taxonomies.

conferences) is the foremost reference, followed by DAFx (ten articles), and ACM and IEEE (nine articles each). Five texts come from books or book chapters. There are also 11 theses/dissertations represented, spread across Ph.D. (six), M.Sc. (four), and B.A. (one). The remaining 36 articles stem from conferences or journals that contributed a single article to the collection.

The yearly spread shown in Fig. 2 shows a steady increase, starting with sporadic articles in 1993, 1995, and 1997 and then moving upward from 1999 to 2017. There is a peak of 12 published articles in 2017, then slightly tapering off after that date. The authors speculate that this decrease may be partly due to growing interest in neural audio synthesis. These are generative approaches, but still not generally controllable with relevant live input (and often not real time), and thus not included in the current review.

3 SOUND TAXONOMY

Based on these procedural audio papers, a concise taxonomy of sounds produced with procedural techniques was then attempted. This was a challenging task because there are some overlaps with every different type of possible groupings. The authors attempted to achieve a relatively even grouping by taking into consideration the sound sources, the physical sounding mechanisms, and various categorizations described in previous articles incorporating taxonomies. Both hierarchical and flat taxonomies influenced the present categorization. Past hierarchical taxonomies have been based on interacting materials [34], states of matter and types of interaction with them [35], as well as detailed sonic feature extraction [36]. Flat taxonomies as in [37] loosely follow Farnell's Practicals section in [10]. The present authors ended up with a flat taxonomy that includes the five large groups listed below.

- Aerodynamic Sounds: Sounds that are generated by the movement of air. These sounds can include the sounds of pneumatic musical instruments as well as swinging and rotating objects.
- Biotic Sounds: Sounds produced by living organisms, such as mammals, birds, and insects. Certain human activities such as vocalizations, crowd applause, and footsteps are included in this group.
- Contact Sounds: Sounds that are generated by the physical interaction of objects. These can include the sounds of impacts, friction, breaking, and other phenomena that result from interactions among objects.
- Machine Sounds: Sounds that are produced by mechanical devices and systems. These can include the sounds of engines, motors, and other machinery that are commonly found in industrial, automotive, and other technical contexts.
- Nature Sounds: Sounds that result from natural elements and processes in the environment. These can include the sounds of water, fire, electricity, and other environmental phenomena.

Purely abstract and electronic sound categories were also initially considered but were eventually discarded. This is mainly because abstract sound models found in the review were too few to consist a full category. However, for both sounds with electronic and abstract elements, it was found that there are some that can fall in previous categories as variations of sound classes with a science fiction aesthetic. Examples are the Lightsaber model found in [38] and the

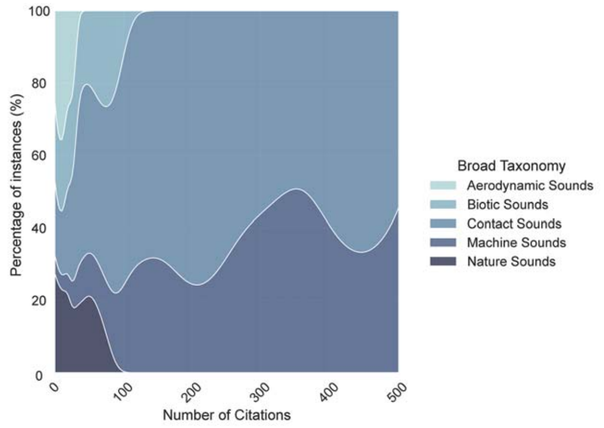

Fig. 3. An analysis of how the number of citations for each paper is spread across the main taxonomy, highlighting what seems to be considered more relevant. Closer to the left one can see the distribution of taxonomic themes along articles that have been cited fewer times. To the right, articles that have been cited over 100 times are mainly distributed between two taxons (Machine and Contact Sounds).

Transporter model found in [10]. Fig. 3 shows an overview of the prevalence of each sound type, when considering how many citations articles working on that taxonomy had.

4 SYNTHESIS TECHNIQUES

Categorizing the various synthesis techniques that have been used for the purposes of procedural audio accurately is a difficult undertaking. This is partly because many researchers in the field use different terminology to describe the same or similar concepts. Techniques that function in a similar way, such as additive and sinusoidal modeling synthesis, were grouped together. However, the authors attempted to not overgeneralize their grouping in order to allow the emergence of deeper insights when analyzing how and when the different techniques were used. Techniques that were very rarely used are grouped in the “Other” category, in an effort to keep the categories balanced and easily understood. The authors ended up with nine distinct synthesis technique groups. Each one of these are outlined in the following subsections.

4.1 Additive Synthesis

Traditionally, additive synthesis was a form of signal-based modeling in which a series of sine waves with independent amplitudes, frequencies, and phases were added together to produce complex waveforms [39]. Additive synthesis has since become the process of modeling sounds as a summation of synthesized audio signals, such as noise signals, sinusoids, and chirp sounds.

The traditional additive synthesis technique was further developed into sinusoidal modeling [40], which represents a signal by a set of sinusoids with time varying parameters. This is shown in Eq. (1):

$$s(t) = \sum_{r=1}^{R} a_r(t)\cos(\theta_r(t)), \tag{1}$$

where $a_r(t)$ and $\theta_r(t)$ are the instantaneous amplitude and phase of the r^{th} sinusoid, respectively, and R is the number of sinusoids. To obtain a sinusoidal representation from an audio sample, an analysis is performed in order to estimate the instantaneous amplitudes and phases of the sinusoids. This estimation is generally achieved by first computing the short time Fourier transform of the sound, then detecting the spectral peaks and measuring their magnitudes, frequencies, and phases. Finally, these are organized as time-varying sinusoidal tracks. The original sound can then be reconstructed using additive synthesis.

This sort of approach is particularly well-suited to tonal sounds [41], such as bells as in [42]. As for controls, duration can be easily modified by allowing control over the time-varying nature of parameters, and frequencies can be scaled to change pitch. Further controls specific to the model, such as how hard a bell is struck, may be devised by mapping a control parameter to relative strength of the sinusoidal components.

4.2 Modal Synthesis

In modal synthesis, sound sources are represented as a collection of resonant vibrating structures, each possessing a number of vibrational modes with nominal frequencies [43]. As noted by Bilbao [44], “modal synthesis may be viewed as a direct physical interpretation of additive synthesis.”

Modal synthesis involves calculating the frequencies generated by the vibrating surface of the object and computing the acoustic radiation caused by each frequency band. Any object that exhibits a few modes and is excited by striking or plucking can be a candidate for modal modeling. To approximate an object’s vibration, a linear vibration equation is used:

$$M\ddot{x} + D\dot{x} + Kx = f, \tag{2}$$

where M, K, and D are mass, stiffness, and damping matrices that depend on the object materials. $f \in \mathbb{R}^{3n}$ is an external force driving the vibration, and $x \in \mathbb{R}^{3n}$ describes the finite element nodal displacement with n nodes. The damping matrix D is usually approximated using the Rayleigh damping model, which represents the damping matrix as a linear combination of mass matrix and stiffness matrix: $D = \alpha M + \beta K$, where the scalars α and β are user-specified parameters.

Linear modal analysis then solves an eigenvalue problem $KU = MUS$ to compute a modal shape matrix U and an eigenvalue matrix S. U describes the vibration pattern of each mode, and S is a diagonal matrix consisting of the squares of the undamped natural frequencies. Eq. (2) can be decoupled into a set of one-dimensional second-order ordinary differential equations, representing the modal vibration of each mode. The *i*th such ordinary differential equation is shown in Eq. (3),

$$\ddot{q}_i + d_i\dot{q}_i + \omega_i^2 q_i = U_i^T f, \tag{3}$$

where U_i is the *i*th column of U, and d_i is the damping parameter of the *i*th mode. The solution to Eq. (3) is a damped sinusoid. This *i*th mode looks like:

$$q_i = a_i e^{-d_i T} \sin(2\pi f_i t + \theta_i), \tag{4}$$

where f_i is the frequency of the mode, d_i the damping coefficient, a_i the excited amplitude, and θi the initial phase. The sum of these modes then gives the generated audio.

Two sets of parameters are of particular importance for achieving desired sound characteristics: vibration frequency ω_i, which determines sound pitch, and damping coefficients d_i, which affect the timbre of materials. For example, a small damping value results in the long ringing sounds that metals often produce, whereas large damping tends to produce more wood-like sounds. In many modal procedural audio models, the user either sets the material, which in turn sets all modal parameters, or controls material parameters (such as density, Young's modulus, or the Coefficient of Restitution), which then maps to relevant changes in specific modal parameters. The resonator-interactor-resonator models described in [24, 45] may all be considered modal synthesis approaches.

4.3 Modulation

Modulation refers to a family of sound synthesis techniques in which some aspect of one signal (the carrier) varies according to an aspect of a second signal (the modulator) [46]. Some of the most common types of modulation synthesis include amplitude modulation, frequency modulation, and ring modulation [47]. In amplitude modulation, the amplitude of a sound wave is modulated by a control signal, resulting in a periodic variation in loudness. In frequency modulation, the frequency of a sound wave is modulated by a control signal, resulting in the creation of complex timbres and harmonics. Ring modulation is a more advanced form of modulation synthesis, in which two sound sources are multiplied and, hence, their frequency components are added and subtracted, creating characteristic timbres and inharmonic distortion.

The basic equation for frequency modulation (FM) synthesis is given by in Eq. (5):

$$x(t) = A_c \sin(\omega_c t + \phi_c + A_m \sin(\omega_m t + \phi_m)). \tag{5}$$

Here, A_c, ω_c, and ϕ_c specify a carrier sound, and A_m, ω_m, and ϕ_m specify a modulator sound. Examples of FM synthesis in procedural audio are the alarm and siren sounds from [10], which aim to replicate the frequency modulation that created those sounds in real-world recordings.

4.4 Physical Modeling Synthesis

Physical modeling refers to a synthesis technique that aims to accurately simulate, using detailed mathematical approximations, the physical processes that give rise to the sound in a real acoustic sound source [48]. The more physics is incorporated into the system, the better the model is considered to be. Physical models are based on fundamental physical properties of a system and its ways of excitation for sound generation. They often involve solving partial differential equations at each time step, as in finite-difference time-domain methods [44], which are known to often produce highly realistic results. However, many physical models are too computationally expensive for real-time use and thus often fall outside the definition of procedural audio. Examples of physical models that are sufficiently lightweight for use in real-time generative audio are the computational fluid dynamics models of [49] and [50].

4.5 Physically Informed Synthesis

Physically informed (also known as physically derived or physically inspired) synthesis is a group of synthesis techniques that draw inspiration from the physical properties of sound phenomena and proceed to model them using less strict approximations than those found in physical modeling techniques. The distinction from physical modeling is that these methods do not fully model the physics behind the sound generation. Instead, they use known high-level physics to guide a signal-based model.

A simple example would be a ball bouncing. In a physical model, the timing of each bounce could be determined based on propagating the ball through a space of coupled regions and taking into account the material to determine dissipation at each impact. In a physically derived model, Newton's equations of motion could give an accurate approximation of the timing of each bounce, and the material's coefficient of restitution would be sufficient to estimate the dissipation at impact.

4.6 Subtractive Synthesis

Subtractive synthesis is a type of sound synthesis in which a rich, harmonically complex sound is filtered and shaped toward some desired goal. Subtractive synthesis typically includes other elements besides just filtering, such as envelope generators to shape the amplitude of the sound over time.

Subtractive synthesis can be considered the complement of additive synthesis, in which simple components are combined to reach a target. Whereas additive synthesis is effective for harmonic sounds, subtractive approaches are often more effective for complex or stochastic sounds.

Subtractive synthesis was the main technique featured in the early music synthesizers [51]. Subtractive synthesis is also a popular approach for generating naturally occurring sound textures such as rain, wind and fire, e.g., [10, 52], because important characteristics of such rich signals can be captured by a carefully filtered noise source.

4.7 Source-Filter Synthesis

This is largely a speech-related synthesis technique that hypothesizes that an acoustic speech signal can be seen as a source signal (the glottal source, or noise generated at a constriction in the vocal tract), filtered with the resonances in the cavities of the vocal tract downstream from the glottis or the constriction [53]. See [54] and references therein for a good overview of source-filter synthesis.

In procedural audio, this synthesis method is effective for synthesizing non-speech vocal sounds, such as a lion roaring or cow mooing [55]. It has also found applications beyond vocalization synthesis, with examples falling in the "Contact" and "Machine" categories in the sound taxonomy. In non-vocal applications, the source is often a rich, broadband signal, and the filter is a model of the processing on that signal because of an interaction or propagation. In much the same way as modal synthesis may be considered a physically derived version of additive synthesis, source-filter synthesis could be considered a physically derived version of subtractive synthesis.

4.8 Waveguide Synthesis

Digital waveguide synthesis models are used to simulate the behavior of wave propagation in physical systems such as strings, tubes, and membranes [56]. They use a combination of delay lines, digital filters, and nonlinear elements to mimic the behavior of the sound source and reflect the geometry and physical properties of the desired instrument. They are efficient for simulating nearly lossless distributed wave media, in which losses and dispersion are consolidated at sparse points along each waveguide.

The Karplus-Strong algorithm [57, 58] simulates a plucked string and is one of the earliest and simplest examples of a waveguide. The basic model uses a delay line of D samples, followed by a low-pass filter, and the feedback of the output back into the delay line. A short noise burst is provided as input to the delay line. Its spectrum then decays to a sine wave at a rate proportional to the length of the delay line. The length of the delay line determines the fundamental frequency of the note that results,

$$f_0 = f_s/D, \tag{6}$$

where f_s is the sampling rate, and D is the length of the delay line in samples.

If the parameters of the low-pass filter are known, the decay rate of the note can also be determined. Suppose that the filter is just a simple gain attenuation. After the initial noise burst, the feedback loop is just doing

$$y[n] = ay[n - D], \tag{7}$$

where y is the output, a is the attenuation in the feedback loop, n is the sample number, and D is the sample delay as before. For a note to decay by 60 dB from its original value, one needs $a^m = 10^{-60/20} = 0.001$ for some number of samples m. This occurs after time

$$t_{60} = \frac{D\ln(0.001)}{\ln(a)f_s}. \tag{8}$$

Further derivation can be used to uncover more characteristics of the generated signal, and extensions to the model can allow for more accurate modeling of plucked string characteristics, such as pickup position [59].

In [60], Karjalainen et al. put the relationship between the Karplus-Strong algorithm and more general digital waveguide models on a firm mathematical basis. This deep understanding is key to procedural audio implementations, because it facilitates implementation of accurate physical modeling of digital waveguides with simpler and often more efficient extensions of the Karplus-Strong algorithm.

4.9 Other Types of Synthesis

This category includes procedural audio models that have not been frequently used and cannot accurately fit into any other category. Statistical models [61, 62], wavelet synthesis examples [63], gestural synthesis examples [9, 64], iterated nonlinear functions [65], and standalone models with unconventional characteristics [10] are the main techniques in this category. A practical example from the aforementioned techniques is the synthesis of continuous (e.g., rain) or impulse-based (e.g., footsteps) stochastic sounds using parameterized wavelet decomposition coefficients and inverse discrete wavelet transform.

Some of the references listed in this category have also been listed in one or more other categories. This is because a reference might define different techniques of producing the same sound and only some of those cannot fit into any other category in Table 1.

5 OVERVIEW OF PROCEDURAL AUDIO

Table 1 summarizes the state of the art by giving a detailed overview of the types of sounds that have been procedurally generated and the techniques that were applied to generate them.

The categories are arguable and were chosen partly to give a balanced table. Nevertheless, one can see that physically informed approaches are very popular, possibly because this is applicable to a very wide variety of sounds and often lends itself well to real-time, controllable implementations. One can also see that certain sounds have been procedurally generated in a large number of papers. Such sounds are either well-known natural sound textures, like wind and fire; common components of sound effects, like rolling; or very popular sound effects in creative content, like footsteps.

One can also observe that certain synthesis types are associated with certain types of sounds. Modal synthesis is popular for generating impact sounds, which is understandable because impacts typically generate strong modal components. Subtractive synthesis features heavily with nature sounds, which are often very rich in spectral content. Aerodynamic sounds are often physically modeled, although this could be largely because of the work of Selfridge, who focused on physical modeling approaches for real-time generation of such sounds [66].

It is intriguing that for popular individual sounds, and for sound classes overall, one often finds that almost every synthesis method has been applied. That is, the community is open to attempting procedural audio using a wide variety of methods for almost any type of sound. Finally, note that there are very few approaches in the "Other" category, suggesting that abstract methods such as those based on

Table 1. A summary of the state of the art in the field of procedural audio. Numbers in each cell correspond to reviewed references in which a sound class is connected with a specific synthesis technique. n = x unique occurrences in the bibliography are provided for both the broad and narrow taxonomies and for the synthesis techniques.

PROCEDURAL AUDIO FIELD SUMMARY										
		SYNTHESIS TYPE								
TAXONOMY	SOUND	Additive (n = 16)	Modal (n = 26)	Modulation (n = 5)	Physical Modelling (n = 22)	Physically Informed (n = 24)	Source-Filter (n = 5)	Subtractive (n = 23)	Waveguide (n = 7)	Other (n = 7)
AERODYNAMIC SOUNDS (n = 13)	Aeolian Harp (n = 3)				[66, 67, 68]					
	Aircraft (n=1)							[69]		
	Airflow (n=1)					[44]				
	Cavity Tone (n=2)				[68, 70]					
	Edge Tone (n=2)				[68, 71]					
	Lightsaber (n=1)							[37]		
	Propeller (n=3)				[66, 68, 72]					
	Rotating Air Fan (n=3)	[10]		[10]		[10]				
	Swinging Objects (n=7)		[18]		[17, 66, 68, 73, 74]			[18]		
BIOTIC SOUNDS (n=22)	Applause (n=4)	[110, 78]				[110, 78]		[126, 79]		
	Bird Call (n=6)	[127]		[10]	[75, 128]			[10]	[10, 76, 77]	
	Drinking Lemonade with a Straw (n=1)					[9]				
	Flocks/Hordes/Swarms (n=2)	[32, 110]				[110]		[32]		
	Footsteps (n=9)	[7, 110]	[2]	[7, 80]		[7, 9, 10, 110, 81, 82]		[7, 80]	[7, 10]	[63]
	Heartbeat (n=1)					[111]				
	Insects (n=2)		[10]	[10]		[112]		[10]		
	Laughter (n=1)							[126]		
	Mammalian Vocalisations (n=5)			[10]	[83]	[10, 55]	[10]		[54]	[61, 64]
	Yelling (n=1)							[126]		
CONTACT SOUNDS (n=33)	Bell (n=5)		[41, 84, 85]		[121]	[111]				
	Bouncing (n=7)	[125]	[9, 125, 113]	[10, 86]	[87]	[9, 114]		[125, 86]		
	Collision (n=15)	[110, 130]	[84, 85, 130, 88, 89, 90, 173, 91]		[49, 87, 91]	[9, 114, 115]		[126]		
	Creaking Door (n=3)	[92]		[131]		[10]		[131]		
	Crumpling (n=3)		[9]			[81, 114]				[9]
	Friction (n=12)	[125, 92]	[84, 85, 125, 93, 94]	[86]		[9, 26, 44, 116]	[29, 93]	[26, 93]		
	Golf Putt (n=1)	[129]							[129]	
	Jackhammer (n=1)					[31]			[31]	
	Near-Rigid Thin Shell Vibrations (n=2)		[95, 122]							
	Non-Rigid Bodies (n=2)			[86]				[86]		
	Pouring Liquid (n=1)		[10]	[10]		[10]				
	Rolling (n=13)	[123]	[9, 28, 93, 94, 123, 62, 174, 96]		[97]	[114, 116, 117]	[28, 93, 123, 62, 96]	[10]		[9, 62]
	Ruler Twang (n=2)		[10, 90]							
	Shuffling Cards (n=1)									[63]
	Soda Bottle Fizz (n=1)							[126]		

Table 1. (*Continued*)

PROCEDURAL AUDIO FIELD SUMMARY										
		SYNTHESIS TYPE								
TAXONOMY	SOUND	Additive (n = 16)	Modal (n = 26)	Modulation (n = 5)	Physical Modelling (n = 22)	Physically Informed (n = 24)	Source-Filter (n = 5)	Subtractive (n = 23)	Waveguide (n = 7)	Other (n = 7)
	Sword Clang (n=2)		[1]					[1, 37]		
MACHINE SOUNDS (n=14)	Alarm (n=2)			[10]				[132]		
	Electric Motor (n=5)	[10]		[10]		[10, 44, 98, 99]				[63]
	Geiger Counter (n=1)							[52]		
	Helicopter (n=1)	[10]	[10]						[10]	
	Industrial Machinery (n=2)		[133]	[86]				[86]		
	Pedestrian Beeping Tone (n=1)							[10]		
	Phone Receiver and DTMF Tones (n=1)	[10]								
	R2D2 (n=1)			[10]						
	Siren (n=1)			[10]		[10]				[10]
	Switch Click (n=1)		[10]	[10]					[10]	
	Telephone Ringing (n=1)	[10]							[10]	
	Ticking Clock (n=1)	[10]	[10]	[10]					[10]	
	Traffic Noise (n=2)	[26]				[26, 100]		[26, 100]	[100]	
	Transporter (n=2)			[10]						[10]
	Vehicle Engine (n=10)	[10, 100, 101]		[10]	[26]	[10, 26, 30, 44, 100, 175]	[123]		[3, 10, 100]	[63]
	Weapons (n=3)		[10]	[10]	[102]	[10]		[124]		
NATURE SOUNDS (n=17)	Avalanche (n=1)		[84]							
	Bubbles (n=3)	[10]				[10, 44, 118]				
	Electricity (n=2)	[10]	[10]			[10, 111]				
	Explosion (n=4)	[10, 130]	[10, 130]		[49]	[10, 44]				
	Fire (n=13)	[10, 118, 103]	[103, 104, 105, 106]	[80, 118]	[49]	[29, 118, 106, 119]		[10, 29, 51, 80, 118, 134]		
	Fluid Events (n=7)	[130]	[9, 130]	[10, 80]	[48]	[107]		[80]		[9, 63]
	Rain (n=9)	[103]	[103, 106]		[108, 109]	[9, 29, 106]		[10, 29, 51]	[10]	[63]
	Room Tone (n=1)					[111]				
	Storm (n=1)									[65]
	Thunder (n=2)	[10]				[10, 29]		[29]		
	Turbulence (n=1)									[65]
	Volcanic Activity (n=1)							[51]		
	Wave (n=4)	[103]	[103]					[37, 51, 120]		
	Wind (n=12)	[118, 103]	[103, 106]	[10]	[73]	[29, 111, 118, 106]		[1, 10, 27, 29, 37, 51, 118, 134]		

wavelets or iterative functions have yet to be picked up as useful tools for the procedural audio community.

6 DESIGN METHODOLOGY

In designing a procedural audio model, the initial step is to establish a design process that informs the synthesis technique and the parameters to be used. Analysis of the papers mentioned in Table 1 can also give further insight into these design methodologies. The current work also attempts to act as a go-to guide for researchers in the field coming up with new projects. The authors have thus included not only the design methodologies that the articles in the authors' list point to, but some methodologies that have been suggested in the literature, but cannot be found in practice, as future work may benefit from considering them, in the following list:

- Physically Derived Equations: The main method of inquiry into how a sound is produced is the construction of a theoretical physical model, with equations coming from classical physics examples. All examples of physical modeling use this strategy by default, but instances of it have been found in 57 of the articles looked at [2, 3, 9, 10, 18, 19, 24, 49, 50, 62, 66–74, 75, 76, 77, 78, 79, 80–82, 83, 55, 84, 85, 86, 87, 88–90, 91, 92, 93, 94, 27, 32, 95, 96, 97, 98, 99, 100, 101, 102, 103–106, 107–109].
- Physically Informed Design: In contrast with the previous methodology, Physically Informed Design is inspired by physics equations but will make generalizations, such as using filtered noise instead of modeling complex phenomena that produce rich, noise-like sounds. It provides simplifications that allow for the use of techniques such as subtractive synthesis and creates faster, computationally cheap models. There were 18 articles that used this approach [9, 10, 45, 65, 110, 33, 111, 112, 113, 114, 115, 116, 117, 31, 118, 105, 119, 65, 120].
- Analytical Recordings: Many studies mention the practice of recording and listening to samples of sounds under analysis, eventually including informal comparative analysis of model versus real sounds and occasionally mentioning strategies such as using low-speed playback [38] for a more-detailed breakdown of the constituent elements of a sound. There were 19 mentions of this approach found [2, 10, 42, 50, 61, 63, 38, 76, 78, 79, 81, 82, 121, 87, 88, 95–123, 124].
- Spectral Analysis: Once a collection of recordings has been obtained, a useful design tool is a spectrum/spectrogram/sonogram that breaks down the frequency components of a given sound, also allowing for a simple visual modal analysis. There were 13 instances of this occurring in the collection [1, 10, 24, 75, 81, 125–87, 89, 122, 103, 107].
- Waveform Analysis: Looking at waveform graphical renditions of recordings will give information on aspects such as the envelope of a sound. It may be a poorer strategy than the previous one, and it accounts for three instances of the works analyzed [24, 38, 103].
- Example-Based Feature Extraction: A larger collection of recordings is sometimes used to extract a set of features that will inform the building of the model, such as statistical moments analysis or any set of features typically associated with machine learning methods. Four instances of this happening have been found [42, 78, 126, 127].
- Physical Deconstruction and Analysis: When one has access to the physical object that produces the sound of interest, breaking it down to its components and analyzing each individually is a strategy akin to modular synthesis, allowing for an isolation of components, such as oscillator, filter and resonance box. Mention of this approach has been found in five articles [3, 98, 99, 100, 101].
- Operational Analysis: This is an alternative to the previous approach that does not actually require the physical break-down of an object, but rather a conceptual/theoretical separation into parts that are known. Examples of this are source-filter or formant analysis, done in general terms. Unlike Physically Derived Equations above, it does not require a full physical/mathematical model to be built. It accounts for four instances of the works under analysis [10, 128, 101, 102].
- Schematics: A non-intrusive way of looking at the physical object features is the use of blueprints or schematics, when available (whenever dealing with the sonic virtualization of built objects that have some physical or electronic document explaining them in detail). Examples of this approach have been found in [10] and [27].
- Impulses and Test Excitations: Mentioned as a possible methodology in [10], recording impulse responses of a physical object of interest is a variation of some of the analytical tools above that can reveal more generic features of a sound. Although this approach may be particularly useful if the system is linear or close to linear, the authors have not actually found it in practical examples in literature.
- Knowledge Engineering: Some projects used the opinion of experts, either individually or in a poll, as an accessory method to design more effective models. This can take the form of fabrication experts or acoustical consultants. This methodology was found to be used in [125] and [129].
- Subjective Evaluation: Although not typical of the design phase, [93] used subjective evaluation to understand the parameters that best fit the perception of rolling sounds, prior to building the final model.
- Heuristics: A large-umbrella methodology for some strategies that do not fit the former, and usually provide refinements to a model that was built with one of the previous approaches. Its inclusion here goes to say that approaches are not static and deviations from the model may create significant improvements. This

is a strategy encountered in 23 articles [1, 7, 28, 52, 64, 65, 38, 127, 128, 110, 79, 80, 130, 131, 29, 129, 132, 99–100, 104, 134, 120].
- Literature: Although it is not a methodology per se, it is worth mentioning that some of the articles under review did not dedicate resources to designing the model and just referenced a work in the theoretical acoustics literature in which a mathematical/physical model had already been constructed. This is something that has been found in ten articles [10, 33, 130, 90, 29, 132, 103, 30, 134, 120].

Fig. 4 shows how these design strategies relate to the type of synthesis used, showing numerically and as a heatmap the amount of cases in which both overlap. Very often works will employ more than one design strategy and synthesis type so there is a greater number of overall instances than the number of papers analyzed. Fig. 5 shows how the design strategies relate to the broader taxonomy of the sounds modeled in each article, showing the overall amount of overlap with size and color. Although there is no clear pattern that emerges from these, one can still make some inferences. The use of recorded examples is ubiquitous across sound types and synthesis types, and the main factor behind recording seems to be the availability of the source (do bear in mind that "Spectral Analysis" also implies recording, and "Analytical Recording" is considered to happen only when analysis is subjective). Physically Derived Equations and Physically Informed Design, as methodologies, are closely tied to Physical Modeling and Physically Informed Synthesis, because these are two cases in which design and process are tightly coupled.

Physically Informed Synthesis will often start from physical equations, which explains why both methodologies are so prevalent for this synthesis type. It is interesting to note that a very high proportion of the modal algorithms (41 out of 57) start from physical equations as well. Subtractive synthesis depends heavily on heuristics and is also high on the use of spectral analysis.

7 EVALUATION

In 1995, Jaffe [135] presented ten different methods for evaluation of synthesis techniques. This framework was used in [136] for an overall evaluation of early synthesis approaches. Five of the evaluation criteria were based on parameter control, three on computational aspects, and only two related to sonic aspects of the synthesis method. Notably, Jaffe's criteria did not include measures of perceptual quality. The assumption was that the techniques were focused on generating interesting sounds, but not necessarily realistic sounds or sounds matching a target.

Despite this early emphasis, evaluation in the field of procedural audio has been severely limited [137]. In a review of 94 published papers on sound texture synthesis [12], only seven contained any perceptual evaluation of the synthesis method. However, this was published in 2011, so an up-to-date analysis is needed.

One of the key aims of procedural audio is to produce a realistic sound, with the added ability to control or interact with the sound [15, 19]. Here, the authors review evaluation approaches for these two main relevant aspects: the perceptual quality, using both objective and subjective approaches, and the controllability of the models. Evaluation approaches that have been applied generally to sound synthesis are considered, as long as they could be applied specifically to procedural audio. The evaluation that has been undertaken in procedural audio models is then classified, and trends toward more and specific types of evaluation are identified.

7.1 Objective Evaluation

There are a range of methods for objective evaluation of synthesized sound effects. However, objective evaluation is often not performed, and there is little to no consistency on which metrics to use. Horner and Wun [138] objectively compared different wavetable synthesis methods using *Relative Spectral Error*, with no comparison to samples. In contrast, [98] compared a synthesis method to reference samples, through visual inspection of spectrograms, and comparison of low-level audio features, but no comparison against other synthesis methods was undertaken.

In [139], synthesis parameter selection was evaluated using a range of low-level audio features, such as Fundamental Frequency, Spectral Shape, Envelope Characteristics, and Overall Duration. They also used the discrete cosine transform of the Mel-Frequency Cepstral Coefficients as a measure of how similar the synthesized sound was to a recorded sample.

Evaluation of perceptual similarity of a piano note synthesis method and recorded samples was attempted in [140]. The authors used Perceptual Evaluation of Audio Quality (PEAQ), an algorithm designed for determining the quality of audio compression codecs, to analyze the sound on a sample-by-sample basis to determine any perceptual artifacts. However, the notes will never be exactly the same if played with slightly different attack or at a different sample time, thus resulting in a perceptual difference in which none exists.

Moffat et al. [36] used feature vectors to compare the sonic similarity of different sound effects. Hamadicharef and Ifeachor [141] attempted to evaluate the perceptual similarity of a piano note synthesis method with a sample using the PEAQ algorithm. PEAQ was designed for determining the quality of audio compression codecs and analyzes the sound on a sample-by-sample basis to determine any perceptual artifacts. This work was further developed by use PEAQ to select parameters for a piano synthesizer to replicate an input audio signal [140]. But the notes will never be exactly the same if played with slightly different attack or at a different sample time, thus resulting in a perceptual difference that should not be attributed to the synthesis model.

[73] and [66] built procedural models of aerodynamic sounds in which estimated and measured physical properties were compared. The output time domain and spectro-

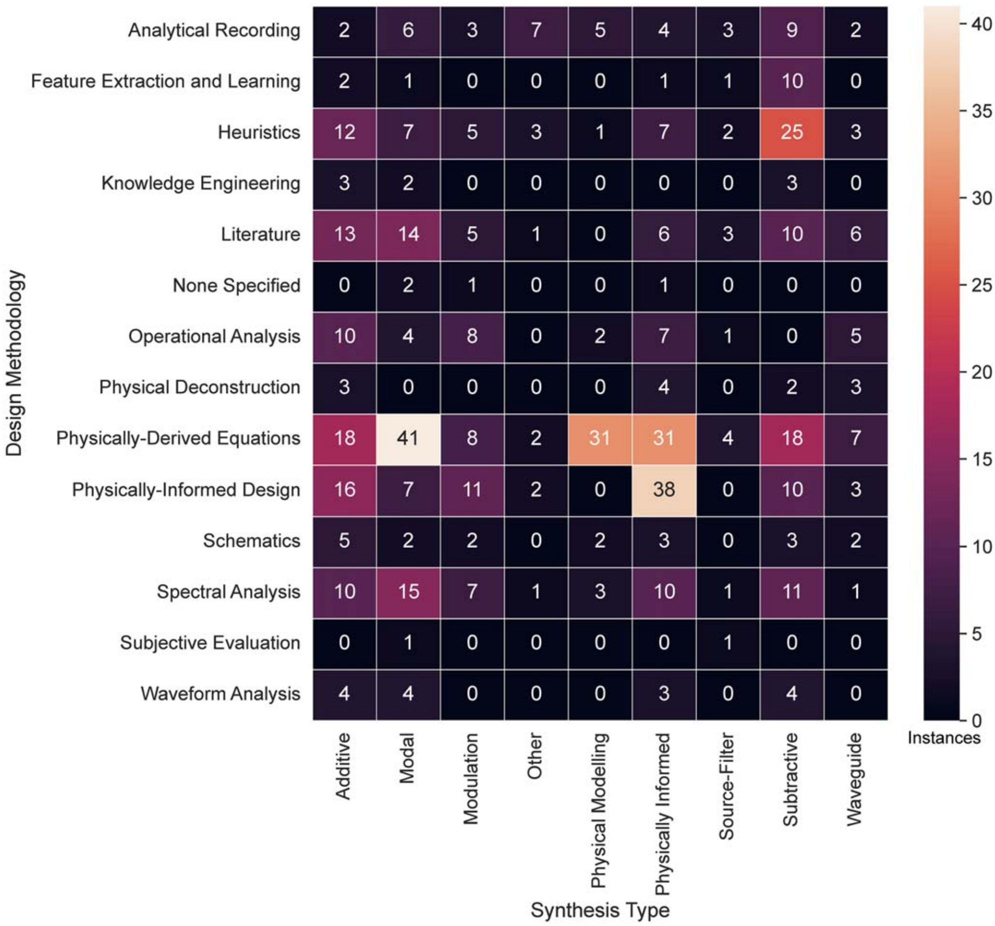

Fig. 4. Relationship between the design strategies used to build the model (*y* axis) and the synthesis type used (*x* axis) for the articles in Table 1. Several articles will make use of more than one methodology and synthesis type; therefore, this heatmap does not have a one-to-one relationship with the articles (i.e., there are many more items here than the total number of articles analyzed).

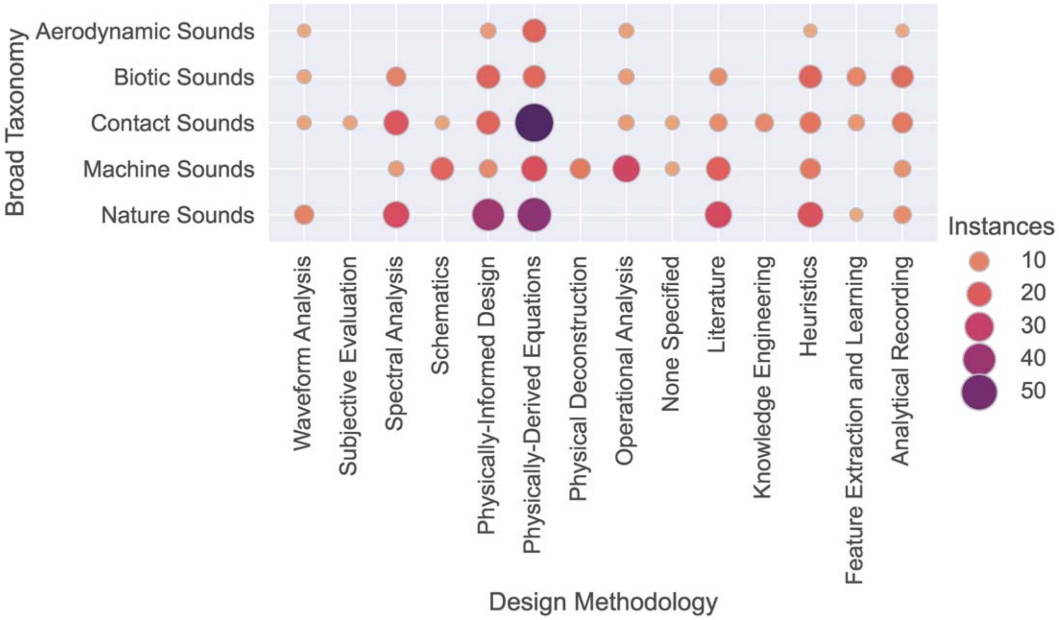

Fig. 5. Relationship between the design strategies used to build the model (*x* axis) and the sound type characterization (*y* axis) for the articles in Table 1.

gram signals were compared visually, including locations of fundamental and harmonics.

7.2 Subjective Evaluation of Sonic Qualities

One of the most important aspects of evaluating a synthesis method is evaluating the perceived quality of the resultant sound. Does the generated audio actually sound as intended? If the desired sound cannot be created to a high perceptual accuracy, then no quantity of sound interaction will make a synthesis model effective.

However, there is no consistently used standard process for evaluating the perceptual realism of sound synthesis. Indirect measurements of the realism or believability of procedurally generated sounds have been reported in numerous studies. Rocchesso et al. [24] reported the effectiveness of procedurally generated sounds as auditory cues to identify aspects of pouring water. The effect of procedural audio approaches on human perception of timbres was assessed in [142, 143]. Performance of procedural audio in sound classification and sound identification tasks was considered in [144, 36, 145] and [146, 147], respectively.

Bonebright et al. [148] discussed three different methods for determining perceptual qualities of audio: identification testing, context-based rating, or attribute rating. They noted that context rating was most appropriate for synthesis of games audio, whereas [142] argued that attribute rating was most appropriate for synthesis of abstract sounds.

McDermott and Simoncelli [146] ran a series of perceptual tests, in which users were asked to perform an identification task in which they needed to pick the right description of the sound from a set of five words. They then performed a pairwise comparison, for which an original sound was played and users had to select which of two options sounded most like the reference, and both options were different synthesized sounds. Participants were then asked to rate the *realism* on a scale of 1–7 for a range of different synthesized and recorded sounds. No formal anchors were identified as the lower bounds for the sound quality.

Gabrielli et al. [144] proposed a form of evaluation they called an *RS Test*, during which participants were played many sound samples only once and each time had to determine whether it was real or synthetic. Hahn also evaluated musical instrument sounds using the RS test [149]. To evaluate instrument synthesis, Scavone et al. [150] created a program for presenting sound effects on a two-dimensional plane using multi-dimensional scaling, and Lakatos et al. [151] asked participants whether they could identify the material dimensions of an impact sound in a two-alternative forced choice experiment. Participants were played samples and had to write free text responses in work by [152].

A different two-dimensional approach was taken in [153]. Here, participants were asked to rate each stimulus in terms of *naturalness* and *concreteness*, for which "An audiovisual is natural if the sound-image ensemble is a credible human action ... [and] ... concrete if it is capable of evoking a physical cause." These two dimensions were then used to compare vocal sketches (using voice to portray and imitate nonspeech sounds), concatenative synthesis, and physical models. However, this paradigm was specific to the problem of vocal sketching. It would need adaptation if used in other contexts, such as evaluation of procedural audio models in animation.

Modern subjective evaluation of procedural audio often takes the form of a multistimulus test, similar to the MUltiple Stimulus Hidden Reference and Anchor (ITU-R BS.1534-3) standard, in which multiple stimuli can be compared against each other.

An evaluation of concatenative synthesis methods was performed via an online test in which participants rated the quality of samples and similarity to the reference sample [154]. There was no randomization of sample order, so potential ordering bias may be an issue, and no recording of the participants' listening conditions was made. They concluded that all concatenative synthesis methods are indistinguishable from each other, in terms of realism and perceived quality. A similar evaluation methodology was undertaken by Mengual et al. [124], in which different synthetic weapon sounds were evaluated with order randomization to remove bias and performed in controlled listening conditions. The conclusion was that modal sounds were synthesized convincingly, but broadband sounds needed further work to improve.

An attribute test was performed in [123], in which participants were asked to rate the quality of "rollingness" of various synthesized rolling sounds but no alternative synthesis methods, samples, or hidden anchors were provided for comparison. In [9], participants were asked to browse through a range of synthesized sounds to find their preferred sound and then asked to rate perceived realism on a seven-point Likert scale. *Perceived realism* was also the evaluation criterion in [19] and [146].

Comparison of synthesis methods, recorded samples, and a specific anchor was undertaken for a range of procedural audio models for aeroacoustic sounds in a series of papers by Selfridge et al. [72, 74, 18, 67, 66]. Comparison of synthesis methods was also used in [155] and in [79] but without the anchor. In [137], it was applied on a large scale to compare synthesis methods against each other for a wide range of sounds. It was further used in [156] to show that no procedural model for thunder performed at all close to a real recording, and this led to an improved thunder model with better performance [157].

7.3 Evaluation of Control and Interaction

Evaluating the control and interaction of a procedural audio model is a vital aspect of understanding how it can and would be used. However, in most cases, the physical interaction that creates the sound will not be suitable for directly driving the individual synthesis parameters, and as such, some mapping layer for parameters and physical properties of a game will often be required [131, 158]. Hoffman and Cook [145] discussed the generalized process of synthesis parameter mapping to perceptual controls through feature vector mapping, and other methods for performing the mapping were given in [143].

User experience tests, in which participants interact with a procedural model through some mapping layer, can be performed to evaluate a series of criteria. Based on research presented in [131, 158, 15], key questions concerning control systems for procedural audio are

- Range: Do the controls allow the user to generate a broad range of sounds, without being too many or overly complex?
- Intuitive: How intuitive and interpretable are the controls, can a user easily find the exact sound they want?
- Perceptible: How much can someone perceive the impact each control makes, at all times, so as to understand what each control does?
- Consistency: Do controls allow for consistent reproduction of sound?
- Reactiveness: Do controls immediately change the sound output, or is there a delay on control parameters, that may impact the ease of usability? Latency up to about 20 ms is often acceptable, so long as the latency is consistent [159].

As part of the Sounding Object project, a large body of work was undertaken focusing on interactions with procedural audio models [24]. Giordano et al. [160] provided a long list of design questions for sounding objects that also suggest ways in which they should be evaluated. Evaluation of the perceived quality of an interaction with a procedural audio model was performed by Böttcher and Serafin [19] and further developed in [161]. In many cases, the control evaluation has to be designed bespoke to the synthesis methods and parametric controls, e.g., [161, 2]. The effectiveness of, or preference for, procedural models when used in a task was evaluated in [24, 74, 162], thus fusing evaluation of realism and control into an overall assessment of user experience.

7.4 Trends in Evaluation

The 99 papers that contained procedural audio models mentioned in Table 1 can be classified as shown in Table 2.

Table 2. State of the art in procedural audio evaluation. The references from Table 1 are classified by evaluation type. The last row gives papers from either subjective evaluation category (comparative or noncomparative) that also contain objective evaluation. The last column gives the average publication year of all papers in a category.

Evaluation Type	Number of Papers	Average Year Published
None	45	2010.24
Proto-Evaluation	15	2009.53
Objective Only	10	2010.00
Subjective, Noncomparative	11	2014.64
Subjective, Comparative	18	2015.67
Total	**99**	**2011.59**
Objective and Subjective	8	2015.75

This shows a marked improvement over the less than 10% of sound texture synthesis papers containing subjective evaluation in [12]. However, there are still significant issues. Forty-five of these papers contained no evaluation, and 15 contained only *proto-evaluation* such as just measurement of computational cost or mentioned evaluation but did not report results. Thus, just over 60% of papers did not report any evaluation giving insight into the effectiveness or characteristics of the model.

Eleven papers reported only noncomparative subjective evaluation (e.g., rating sound quality of just the proposed model or trying to identify whether stimuli are from a real recording or from just that model), as opposed to comparative subjective evaluation, in which their procedural audio approach was compared against other methods to generate the same type of sound. Thus over a third of papers with subjective evaluation did not provide a means by which the evaluation could be put in context. For example, "seven out of ten participants agreed that this sounded realistic" only becomes meaningful once one knows that other methods are considered far less or far more realistic.

Looking at the average publication dates of each category, one can also see an upward trend. Papers without subjective evaluation had an average publication date around 2010, whereas those with reported subjective evaluation results had an average publication date around 2015. This trend toward more evaluation suggests that it may become easier to identify which procedural audio methods are effective for which sounds.

8 EMERGING FIELDS IN PROCEDURAL AUDIO

An area of procedural audio that has seen significant advancements in recent years is the use of machine learning and artificial intelligence. Machine learning algorithms can be used to analyze existing audio assets, allowing the creation of new, similar content based on the existing data [163, 155]. Alternatively, machine learning algorithms can be used to control parameters of sound synthesis models for the automatic real-time creation of different types of sounds [164]. Both approaches can be useful in creating variations of sound effects and music, adding richness and reducing repetition in the audio experience.

Neural audio synthesis (NAS) is an emerging subfield of machine learning in which multilayer neural networks are used to generate sounds [21]. They have the advantage that they generally result in sounds of exceptionally high quality [165], as compared to other generative audio approaches.

In [166], convolutional neural networks were used to create sound textures such as rain, wind, and crowds. However, generative adversarial networks have been shown to achieve very high performance with speech synthesis. Thus, they were employed by [167] for bird sounds, [163] for knocking sounds, and [155] for footsteps.

However, in their current state, NAS approaches would not be considered procedural audio because the control is very limited. For instance, [155] presented a method to generate a single footstep, for different surfaces. The perceived realism was on par with high-quality recordings.

But the NAS approach did not generate continual footsteps, i.e., walking and running. Nor did it provide control for the firmness of the steps or any other aspect that might be a real-time input from a user, game, or simulation environment.

Real-time control may be possible, especially with large and diverse training data. In particular, Large Language Models may be used to provide meaningful semantic controls for neural audio synthesis. However, such approaches are only just beginning to be undertaken. A Max/MSP interface [168] allows RAVE [169] and other NAS systems to be integrated into a Max patch. But existing controls are only for running and configuring the model, such as connecting encoder and decoder methods. And although model behavior could be modified using Max MSP objects, the only examples of this are for abstract controls, such as a temperature slider for controlling the randomness of predictions. Devis et al. [170] tackles the issue of deep generative audio with expressive and continuous descriptor-based control. However, no examples are given for non-speech, non-music signals, and the controls for signal generation are all abstract and not specific to the type of sound being generated, e.g., spectral centroid, RMS, and boominess.

In [21], the authors posed the question of whether physical modeling or machine learning (and NAS specifically) approaches should be used to synthesize musical instrument sounds. However, the two approaches are not mutually exclusive. Training data in a neural audio synthesis technique can use a combination of real recordings and outputs from physical models. And unspecified parameters of a physical model can be determined by optimizing parameters based on analysis of recordings.

Such an approach has been implemented in, for instance, [171], in which an FM synthesizer was made differentiable and embedded in a neural network architecture, so that it could be trained to synthesize musical instrument sounds. Masuda and Saito [172] looked in detail at the idea of NAS being applied to optimize a procedural audio model. In their case, the model was a generic music synthesizer with effect modules and envelope generators, but it could just as easily have been almost any of the models from Table 1. Not only could this be used to optimize parameter settings to match recordings, but investigation of the loss function could provide insight into the quality and usefulness of different procedural audio models.

9 CONCLUSION

The state of the art in procedural audio has shown remarkable progress over the past few decades. Procedural audio techniques have evolved from simple synthesis algorithms to complex systems that use advanced computational methods to create highly interactive, diverse, and realistic soundscapes. The use of procedural audio has significantly impacted various fields, including video games, virtual reality, music, film, and multimedia. By generating sound procedurally, developers and designers can create immersive and dynamic experiences that respond to user input and environmental changes in real time.

There have been impressive advances in procedural audio technology, but there are still significant challenges and limitations to be addressed. For example, creating high-quality procedural audio that accurately mimics the complexity and nuances of real-world sounds can be a difficult task. Additionally, designing procedural audio systems that can be easily integrated into existing software and hardware platforms remains a challenge.

Despite the large body of work on evaluation of sound synthesis, including procedural audio, many of the methods listed in Table 1 have not been evaluated in terms of realism or related attributes. When evaluation has been performed, it is often not subjective, and it is even rarer for it to be comparative, in which the proposed technique is compared against alternatives. Nor have standard methodologies been established. Without understanding of current synthesis techniques, their benefits, and their weaknesses, it is not possible to understand where the current deficits exist. This failing of the sound synthesis community to address evaluation is a clear contributing factor to the lack of understanding of the current state of the art in sound synthesis.

As is evident from the literature, it is never expected that a single synthesis method is effectively able to produce all possible sounds. In every case, there may be a range of synthesis approaches that are appropriate. However, this simply highlights the importance of evaluation. Identification of suitable use cases and occasions in which a particular sound synthesis method is applicable is vital for the adoption of procedural audio. As the demand for more immersive and interactive experiences continues to grow, procedural audio is sure to play an increasingly important role in shaping the future of digital media.

10 ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training in Intelligent Games and Game Intelligence (iGGi) [EP/S022325/1].

11 REFERENCES

[1] J.-L. Sinclair, “Procedural Audio: Beyond Samples,” in *Principles of Game Audio and Sound Design*, pp. 238–250 (Focal Press, New York, NY, 2020). https://www.taylorfrancis.com/books/mono/10.4324/9781315184432/principles-game-audio-sound-design-jean-luc-sinclair.

[2] R. Nordahl, S. Serafin, and L. Turchet, “Sound Synthesis and Evaluation of Interactive Footsteps for Virtual Reality Applications,” in *Proceedings of the IEEE Virtual Reality Conference (VR)*, pp. 147–153 (Boston, MA) (2010 Mar.). https://doi.org/10.1109/VR.2010.5444796.

[3] S. Baldan, H. Lachambre, S. Delle Monache, and P. Boussard, “Physically Informed Car Engine Sound Synthesis for Virtual and Augmented Environments,” in *Proceedings of the IEEE 2nd VR Workshop on Sonic Interactions*

for Virtual Environments (SIVE), pp. 1–6 (Arles, France) (2015 Mar.). https://doi.org/10.1109/SIVE.2015.7361287.

[4] F. Avanzini, "Procedural Modeling of Interactive Sound Sources in Virtual Reality," in M. Geronazzo and S. Serafin (Eds.), *Sonic Interactions in Virtual Environments*, pp. 49–76 (Springer, Cham, Switzerland, 2022). https://doi.org/10.1007/978-3-031-04021-4_2.

[5] J. Rawlinson, "w[i]nd: An Interactive and Generative Audiovisual Composition/Installation in Procedural and Orchestrated Space," in *Proceedings of the 9th Conference on Computation, Communication, Aesthetics & X (xCoAx)*, pp. 483–489 (Online) (2020 Jul.). https://2021.xcoax.org/data/pdf/xCoAx2021-Rawlinson.pdf.

[6] Y. Cao, H. Hou, Y. Liu, L. Tang, and Y. Li, "Engine Order Sound Simulation by Active Sound Generation for Electric Vehicles," *SAE Int. J. Veh. Dyn. Stab. NVH*, vol. 4, no. 2, pp. 151–164 (2020 Feb.). https://doi.org/10.4271/10-04-02-0011.

[7] B. Horsak, R. Dlapka, M. Iber, et al., "SONIGait: A Wireless Instrumented Insole Device for Real-Time Sonification of Gait," *J. Multimodal User Interfaces*, vol. 10, no. 3, pp. 195–206 (2016 Apr.). https://doi.org/10.1007/s12193-016-0216-9.

[8] P. R. Cook, *Real Sound Synthesis for Interactive Applications* (A K Peters/CRC Press, New York, NY, 2003). https://doi.org/10.1201/b19597.

[9] D. Rocchesso and F. Fontana, *The Sounding Object* (Mondo Estremo, Italy, 2003). https://books.google.co.uk/books?id=f7fQ0xfwmE4C.

[10] A. Farnell, *Designing Sound* (MIT Press, Cambridge, MA, 2010). https://mitpress.mit.edu/9780262014410/designing-sound/.

[11] G. Strobl, G. Eckel, and D. Rocchesso, "Sound Texture Modeling: A Survey," in *Proceedings of the Sound Music Computing Conference* (Marseille, France) (2006 May). https://hal.science/hal-03013698/document.

[12] D. Schwarz, "State of the Art in Sound Texture Synthesis," in *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx)*, pp. 221–232 (Paris, France) (2011 Sep.). https://hal.science/hal-01161296/document.

[13] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in Audio Texture Analysis, Synthesis, and Applications," *J. Audio Eng. Soc.*, vol. 70, no. 3, pp. 108–127 (2022 Mar.). https://doi.org/10.17743/jaes.2021.0060.

[14] A. Misra and P. R. Cook, "Toward Synthesized Environments: A Survey of Analysis and Synthesis Methods for Sound Designers and Composers," in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 155–162 (Montreal, Canada) (2009 Aug.). https://hdl.handle.net/2027/spo.bbp2372.2009.035.

[15] D. Moffat, R. Selfridge, and J. D. Reiss, "Sound Effect Synthesis," in M. Filimowicz (Ed.), *Foundations in Sound Design for Interactive Media*, pp. 274–299 (Routledge, New York, NY, 2019). http://doi.org/10.4324/9781315106342-13.

[16] S. A. Rizzi and A. K. Sahai, "Auralization of Air Vehicle Noise for Community Noise Assessment," *CEAS Aeronaut.*, vol. 10, no. 1, pp. 313–334 (2019 Mar.). https://doi.org/10.1007/s13272-019-00373-6.

[17] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization—An Overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861–875 (1993 Nov.). https://www.mattmontag.com/auralization/media/Auralization%20-%20An%20Overview.pdf.

[18] R. Selfridge, D. Moffat, and J. D. Reiss, "Sound Synthesis of Objects Swinging Through Air Using Physical Models," *Appl. Sci.*, vol. 7, no. 11, paper 1177 (2017 Nov.). https://doi.org/10.3390/app7111177.

[19] N. Böttcher and S. Serafin, "Design and Evaluation of Physically Inspired Models of Sound Effects in Computer Games," in *Proceedings of the 35th AES International Conference: Audio for Games* (2009 Feb.), paper 13. https://vbn.aau.dk/ws/portalfiles/portal/19089661/AES_2008_final_noCrash_version.pdf.

[20] F. Su and C. Joslin, "Toward Generating Realistic Sounds for Soft Bodies: A Review," in *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, pp. 199–206 (Nottingham, UK) (2019 Sep.). https://doi.org/10.1145/3356590.3356620.

[21] S. H. Hawley, V. Chatziiannou, and A. Morrison, "Synthesis of Musical Instrument Sounds: Physics-Based Modeling or Machine Learning," *Acoust. Today*, vol. 16, no. 1, pp. 20–28 (2020 Spring). https://doi.org/10.1121/AT.2020.16.1.20.

[22] A. Natsiou and S. O'Leary, "Audio Representations for Deep Learning in Sound Synthesis: A Review," in *Proceedings of the IEEE/ACS 18th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–8 (Tangier, Morocco) (2021 Nov.). https://doi.org/10.1109/AICCSA53542.2021.9686838.

[23] A. Farnell, "An Introduction to Procedural Audio and Its Application in Computer Games," in *Proceedings of the Audio Mostly Conference*, vol. 23, pp. 1–31 (Ilmenau, Germany) (2007 Sep.). https://www.semanticscholar.org/paper/An-introduction-to-procedural-audio-and-its-in-Farnell/10bd6392b1c79667be47b207834af42c56cf87e2.

[24] D. Rocchesso, R. Bresin, and M. Fernstrom, "Sounding Objects," *IEEE Multimed.*, vol. 10, no. 2, pp. 42–52 (2003 Apr.). https://doi.org/10.1109/MMUL.2003.1195160.

[25] C. Roads, "Introduction to Granular Synthesis," *Comput. Music J.*, vol. 12, no. 2, pp. 11–13 (1988 Summer). https://doi.org/10.2307/3679937.

[26] D. Schwarz, "Corpus-Based Concatenative Synthesis," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 92–104 (2007 Mar.). https://doi.org/10.1109/MSP.2007.323274.

[27] Y. Fu, *Auralisation of Traffic Flow using Procedural Audio Methods*, Ph.D. thesis, University of York, York, UK (2021 Jan.). https://etheses.whiterose.ac.uk/28802/.

[28] J. Thorsteinsson, "Real Time Procedural Wind Soundscape: The Effect of Procedural Wind Soundscape on Navigation in Virtual 3D Space," Bachelor's degree project, University of Skövde, Skövde, Sweden (2015 Aug.). https://www.diva-portal.org/smash/get/diva2:849694/FULLTEXT01.pdf.

[29] L. Pruvost, B. Scherrer, M. Aramaki, S. Ystad, and R. Kronland-Martinet, “Perception-Based Interactive Sound Synthesis of Morphing Solids’ Interactions,” in *Proceedings of the SIGGRAPH Asia Conference*, paper 17 (Kobe, Japan) (2015 Nov.). https://doi.org/10.1145/2820903.2820914.

[30] C. Floroiu and I. Gibson, “Towards a User-Friendly System for Environmental Sound Synthesis in Games,” presented at the *150th Convention of the Audio Engineering Society* (2021 May), paper 10466.

[31] M. Luczynski, and S. Brachmanski, “Mathematical Model of the Acoustic Signal Generated by the Combustion Engine,” presented at the *142nd Convention of the Audio Engineering Society* (2017 May), paper 9717.

[32] S. Oksanen, J. Parker, and V. Välimäki, “Physically Informed Synthesis of Jackhammer Tool Impact Sounds,” in *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx)*, pp. 168–171 (Maynooth, Ireland) (2013 Sep.). https://research.aalto.fi/en/publications/physically-informed-synthesis-of-jackhammer-tool-impact-sounds.

[33] J. Pekonen and A. Jylhä, “3-D Sound Synthesis of a Honeybee Swarm,” presented at the *127th Convention of the Audio Engineering Society* (2009 Oct.), paper 7883. https://research.aalto.fi/en/publications/3d-sound-synthesis-of-a-honeybee-swarm.

[34] W. W. Gaver, “What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception,” *Ecol. Psychol.*, vol. 5, no. 1, pp. 1–29 (1993 Mar.). https://doi.org/10.1207/s15326969eco0501_1.

[35] G. Lemaitre and L. M. Heller, “Evidence for a Basic Level in a Taxonomy of Everyday Action Sounds,” *Experiment. Brain Res.*, vol. 226, pp. 253–264 (2013 Feb.). https://doi.org/10.1007/s00221-013-3430-7.

[36] D. Moffat, D. Ronan, J. D. Reiss, et al., “Unsupervised Taxonomy of Sound Effects,” in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx)*, pp. 428–435 (Edinburgh, UK) (2017 Sep.). https://dafx17.eca.ed.ac.uk/papers/DAFx17_paper_26.pdf.

[37] P. Bahadoran, A. Benito, T. Vassallo, and J. D. Reiss, “Fxive: A Web Platform for Procedural Sound Synthesis,” presented at the *144th Convention of the Audio Engineering Society* (2018 May), e-Brief 416.

[38] T. Hillerson, *Programming Sound With Pure Data: Make Your Apps Come Alive With Dynamic Audio* (Pragmatic Bookshelf, Dallas, TX, 2014). https://pragprog.com/titles/thsound/programming-sound-with-pure-data/.

[39] R. C. Maher, “Sinewave Additive Synthesis Revisited,” presented at the *91st Convention of the Audio Engineering Society* (1991 Oct.), paper 3128. https://virus.montana.edu/rmaher/publications/maher_aes_1091_3128.pdf.

[40] X. Serra, *A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic Plus Stochastic Decomposition*, Doctoral thesis, Stanford University, Stanford, CA (1989 Oct.). https://repositori.upf.edu/handle/10230/34072.

[41] J. W. Beauchamp, *Analysis, Synthesis, and Perception of Musical Sounds* (Springer, New York, NY, 2007). https://doi.org/10.1007/978-0-387-32576-7.

[42] E. K. Canfield-Dafilou and K. J. Werner, “Modal Audio Effects: A Carillon Case Study,” in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx)*, pp. 247–54 (Edinburgh, UK) (2017 Sep.). https://www.dafx17.eca.ed.ac.uk/papers/DAFx17_paper_22.pdf.

[43] I. Bisnovatyi, “Flexible Software Framework for Modal Synthesis,” in *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)* (Verona, Italy) (2000 Dec.). https://doi.org/10.48550/arXiv.cs/0103006.

[44] S. Bilbao, *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics* (Wiley, Chichester, UK, 2009). https://doi.org/10.1002/9780470749012.

[45] S. Baldan, S. Delle Monache, and D. Rocchesso, “The Sound Design Toolkit,” *SoftwareX*, vol. 6, pp. 255–260 (2017 Jun.). https://doi.org/10.1016/j.softx.2017.06.003.

[46] C. Roads, *The Computer Music Tutorial* (MIT Press, Cambridge, MA, 1996). https://mitpress.mit.edu/9780262680820/the-computer-music-tutorial/.

[47] M. Russ, *Sound Synthesis and Sampling* (Routledge, New York, NY, 2004), 2nd ed. https://doi.org/10.4324/9780080481043.

[48] J. O.III Smith,, *Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects*, https://ccrma.stanford.edu/~jos/pasp/ (2010 Dec.).

[49] W. Moss, H. Yeh, J.-M. Hong, M. C. Lin, and D. Manocha, “Sounding Liquids: Automatic Sound Synthesis From Fluid Simulation,” *ACM Trans. Graph.*, vol. 29, no. 3, paper 21 (2010 Jul.). https://doi.org/10.1145/1805964.1805965.

[50] Y. Dobashi, T. Yamamoto, T. Nishita, “Synthesizing Sound From Turbulent Field Using Sound Textures for Interactive Fluid Simulation,” *Comput. Graph. Forum*, vol. 23, no. 3, pp. 539–545 (2004 Aug.). https://doi.org/10.1111/j.1467-8659.2004.00785.x.

[51] J. Pekonen and V. Välimäki, “The Brief History of Virtual Analog Synthesis,” in *Proceedings of the 6th Forum Acusticum*, pp. 461–466 (Aalborg, Denmark) (2011 Jun.).

[52] M. Hansen, “Weather Organ: A Sparse Stochastic Synthesizer in Faust,” in *Proceedings of the 1st International Faust Conference (IFC)* (Mainz, Germany) (2018 Jul.). https://www.ifc18.uni-mainz.de/papers/hansen.pdf.

[53] G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, The Netherlands, 1960).

[54] A. M. Taylor and D. Reby, “The Contribution of Source–Filter Theory to Mammal Vocal Communication Research,” *J. Zool.*, vol. 280, no. 3, pp. 221–236 (2010 Feb.). https://doi.org/10.1111/j.1469-7998.2009.00661.x.

[55] W. Wilkinson and J. D. Reiss, “A Synthesis Model for Mammalian Vocalization Sound Effects,” in *Proceedings of the 61st AES International Conference: Audio for Games* (2016 Feb.), paper 4-3. https://www.eecs.qmul.ac.uk/~josh/documents/2016/wilkinson%20reiss%20-%202016.pdf.

[56] J. Smith, "A Basic Introduction to Digital Waveguide Synthesis (for the Technically Inclined)," https://ccrma.stanford.edu/~jos/swgt/ (2006 Feb.).

[57] K. Karplus and A. Strong, "Digital Synthesis of Plucked-String and Drum Timbres," *Comput. Music J.*, vol. 7, no. 2, pp. 43–55 (1983 Summer). https://doi.org/10.2307/3680062.

[58] D. A. Jaffe and J. O. Smith, "Extensions of the Karplus-Strong Plucked-String Algorithm," *Comput. Music J.*, vol. 7, no. 2, pp. 56–69 (1983 Summer). https://doi.org/10.2307/3680063.

[59] C. R. Sullivan, "Extending the Karplus-Strong Algorithm to Synthesize Electric Guitar Timbres With Distortion and Feedback," *Comput. Music J.*, vol. 14, no. 3, pp. 26–37 (1990 Autumn). https://doi.org/10.2307/3679957.

[60] M. Karjalainen, V. Välimäki, and T. Tolonen, "Plucked-String Models: From the Karplus-Strong Algorithm to Digital Waveguides and Beyond," *Comput. Music J.*, vol. 22, no. 3, pp. 17–32 (1998 Autumn). https://doi.org/10.2307/3681155.

[61] R. E. Turner, *Statistical Models for Natural Sounds*, Ph.D. thesis, University College London, London, UK (2010 Jan.). https://discovery.ucl.ac.uk/id/eprint/19231/1/19231.pdf.

[62] J. S. Lee, P. Depalle, and G. Scavone, "Analysis/Synthesis of Rolling Sounds Using a Source-Filter Approach," in *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx)* (Graz, Austria) (2010 Sep.). https://dafx10.iem.at/proceedings/papers/LeeDepalleScavone_DAFx10_P104.pdf.

[63] N. E. Miner and T. P. Caudell, "Using Wavelets to Synthesize Stochastic-Based Sounds for Immersive Virtual Environments," *ACM Trans. Appl. Percept.*, vol. 2, no. 4, pp. 521–528 (2005 Oct.). https://doi.org/10.1145/1101530.1101552.

[64] R. Hamilton, "The Procedural Sounds and Music of ECHO::Canyon," in *Proceedings of the 40th International Computer Music Conference (ICMC)*, pp. 449–455 (Athens, Greece) (2014 Sep.). https://ccrma.stanford.edu/~rob/papers/hamilton_icmc2014.pdf.

[65] A. Di Scipio, "Synthesis of Environmental Sound Textures by Iterated Nonlinear Functions," in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx)*, pp. 109–117 (Trondheim, Normandy) (1999 Dec.). https://www.nic.funet.fi/index/Science/audio/dafx/1999/www.tele.ntnu.no/akustikk/meetings/DAFx99/discipio.pdf.

[66] R. Selfridge, D. Moffat, E. J. Avital, and J. D. Reiss, "Creating Real-Time Aeroacoustic Sound Effects Using Physically Informed Models," *J. Audio Eng. Soc.*, vol. 66, no. 7/8, pp. 594–607 (2018 Aug.). https://doi.org/10.17743/jaes.2018.0033.

[67] R. Selfridge, D. J. Moffat, J. D. Reiss, and E. J. Avital, "Real-Time Physical Model of an Aeolian Harp," in *Proceedings of the 24th International Congress on Sound and Vibration* (London, UK) (2017 Jul.). https://www.eecs.qmul.ac.uk/~josh/documents/2017/Selfridge%20Moffat%20Reiss%20Avital%20-%20ICSV2017.pdf.

[68] R. Selfridge, *Real-Time Sound Synthesis of Aeroacoustic Sounds using Physically Derived Models*, Ph.D. thesis, Queen Mary University of London, London, UK (2019 May). https://qmro.qmul.ac.uk/xmlui/handle/123456789/58445.

[69] J. Périsse, P. Chevret, C. Thirard, R. Maier, and J.-M. Nogues, "Towards Virtual Sound Aircraft Simulation," in *Proceedings of the Forum Acusticum Conference*, pp. 827–832 (Budapest, Hungary) (2005 Jan.).

[70] R. Selfridge, J. D. Reiss, and E. Avital, "Physically Derived Synthesis Model of a Cavity Tone," in *Proceedings of the 20th Digital Audio Effects Conference (DAFx)*, pp. 307–314 (Edinburgh, UK) (2017 Sep.). https://www.dafx17.eca.ed.ac.uk/papers/DAFx17_paper_69.pdf.

[71] R. Selfridge, J. D. Reiss, and E. J. Avital, "Physically Derived Synthesis Model of an Edge Tone," presented at the *144th Convention of the Audio Engineering Society* (2018 May), paper 9956.

[72] R. Selfridge, D. Moffat, and J. D. Reiss, "Physically Derived Sound Synthesis Model of a Propeller," in *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, paper 16 (London, UK) (2017 Aug.). https://doi.org/10.1145/3123514.3123524.

[73] R. Selfridge, J. D. Reiss, E. J. Avital, and X. Tang, "Physically Derived Synthesis Model of an Aeolian Tone," presented at the *141st Convention of the Audio Engineering Society* (2016 Sep.), paper 9679. https://www.eecs.qmul.ac.uk/~josh/documents/2016/Selfridge%20Reiss%20AES141%202016.pdf.

[74] R. Selfridge, D. Moffat, and J. Reiss, "Real-Time Physical Model for Synthesis of Sword Swing Sounds," in *Proceedings of the 14th Sound and Music Computing Conference*, pp. 299–305 (Espoo, Finland) (2017 Jul.).

[75] T. Smyth and J. O. Smith, "The Sounds of the Avian Syrinx—Are They Really Flute-Like?" in *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx)*, pp. 199–202 (Hamburg, Germany) (2002 Sep.). https://musicweb.ucsd.edu/trsmyth/pubs/dafx2002.pdf.

[76] M. Kahrs and F. Avanzini, "Computer Synthesis of Bird Songs and Calls," *J. Acoust. Soc. Am.*, vol. 111, pp. 23–27 (2002 May). https://doi.org/10.1121/1.4778131.

[77] T. Smyth and J. O. Smith, "The Syrinx: Nature's Hybrid Wind Instrument," *J. Acoust. Soc. Am.*, vol. 112, no. 5, p. 2240 (2002 Oct.). https://doi.org/10.1121/1.4778889.

[78] L. Peltola, C. Erkut, P. R. Cook, and V. Valimaki, "Synthesis of Hand Clapping Sounds," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 3, pp. 1021–1029 (2007 Mar.). https://doi.org/10.1109/TASL.2006.885924.

[79] J. R. R. Lee and J. D. Reiss, "Real-Time Sound Synthesis of Audience Applause," *J. Audio Eng. Soc.*, vol. 68, no. 4, pp. 261–272 (2020 Apr.). https://doi.org/10.17743/jaes.2020.0006.

[80] B. del Toro, *Using Procedural Audio to Control an Algorithmic Composition That is Controlled by a Computer Game*, Master's thesis, Florida International University, Miami, FL (2013 Apr.). https://doi.org/10.25148/etd.FI13080501.

[81] F. Fontana and R. Bresin, "Physics-Based Sound Synthesis and Control: Crushing, Walking and Running by Crumpling Sounds," in *Proceedings of the 14th Colloquium on Musical Informatics*, pp. 109–114 (Firenze, Italy) (2003 May). https://www.speech.kth.se/prod/publications/files/981.pdf.

[82] A. J. Farnell, "Marching Onwards: Procedural Synthetic Footsteps for Video Games and Animation," presented at the *Pure Data Convention* (Montreal, Canada) (2007 Aug.).

[83] R. K. Moore, "A Real-Time Parametric General-Purpose Mammalian Vocal Synthesiser," in *Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 2636–2640 (San Francisco, CA) (2016 Sep.). https://doi.org/10.21437/Interspeech.2016-841.

[84] C. P. van den Doel, *Sound Synthesis for Virtual Reality and Computer Games*, Ph.D. thesis, University of British Columbia, Vancouver, Canada (1998 Nov.). https://doi.org/10.14288/1.0051161.

[85] K. van den Doel and D. K. Pai, "Modal Synthesis for Vibrating Objects," in K. Greenebaum and R. Barzel (Eds.), *Audio Anecdotes III*, pp. 99–120 (A K Peters/CRC Press, New York, NY, 2007). https://doi.org/10.1201/9781439864869-7.

[86] W. W. Gaver, "Synthesizing Auditory Icons," in *Proceedings of the INTERACT and CHI Conference on Human Factors in Computing Systems*, pp. 228–235 (Amsterdam, The Netherlands) (1993 Apr.). https://doi.org/10.1145/169059.169184.

[87] J. F. O'Brien, P. R. Cook, and G. Essl, "Synthesizing Sounds From Physically Based Motion," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 529–536 (Los Angeles, CA) (2001 Aug.). https://doi.org/10.1145/383259.383321.

[88] W. J. Wilkinson, J. D. Reiss, and D. Stowell, "Latent Force Models for Sound: Learning Modal Synthesis Parameters and Excitation Functions From Audio Recordings," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx)*, pp. 56–63 (Edinburgh, UK) (2017 Sep.). https://qmro.qmul.ac.uk/xmlui/handle/123456789/28403.

[89] J. F. O'Brien, C. Shen, and C. M. Gatchalian, "Synthesizing Sounds From Rigid-Body Simulations," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 175–181 (San Antonio, TX) (2002 Jul.). https://doi.org/10.1145/545261.545290.

[90] C. Zheng and D. L. James, "Toward High-Quality Modal Contact Sound," in *Proceedings of the ACM SIGGRAPH Conference*, paper 38 (Vancouver, Canada) (2011 Aug.). https://doi.org/10.1145/1964921.1964933.

[91] J.-H. Wang, *Physics-Based Sound Synthesis Using Time-Domain Methods*, Ph.D. thesis, Stanford University, Stanford, CA (2019 Dec.). https://purl.stanford.edu/sk366cr9606.

[92] P. Sanchez and J. D. Reiss, "Real-Time Synthesis of Sound Effects Caused by the Interaction Between Two Solids," presented at the *146th Convention of the Audio Engineering Society* (2019 Mar.), paper 10184. https://www.eecs.qmul.ac.uk/∼josh/documents/2019/20317.pdf.

[93] S. Conan, E. Thoret, M. Aramaki, et al., "An Intuitive Synthesizer of Continuous-Interaction Sounds: Rubbing, Scratching, and Rolling," *Comput. Music J.*, vol. 38, no. 4, pp. 24–37 (2014 Dec.). https://doi.org/10.1162/COMJ_a_00266.

[94] K. Van Den Doel, P. G. Kry, and D. K. Pai, "FoleyAutomatic: Physically-Based Sound Effects for Interactive Simulation and Animation," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 537–544 (Los Angeles, CA) (2001 Aug.). https://doi.org/10.1145/383259.383322.

[95] J. N. Chadwick, S. S. An, and D. L. James, "Harmonic Shells: A Practical Nonlinear Sound Model for Near-Rigid Thin Shells," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–10 (2009 Dec.). https://doi.org/10.1145/1618452.1618465.

[96] S. Conan, O. Derrien, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "A Synthesis Model With Intuitive Control Capabilities for Rolling Sounds," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 8, pp. 1260–1273 (2014 May). https://doi.org/10.1109/TASLP.2014.2327297.

[97] N. Raghuvanshi and M. C. Lin, "Interactive Sound Synthesis for Large Scale Environments," in *Proceedings of the Symposium on Interactive 3D Graphics and Games*, pp. 101–108 (Redwood City, CA) (2006 Mar.). https://doi.org/10.1145/1111411.1111429.

[98] S. Hendry and J. D. Reiss, "Physical Modeling and Synthesis of Motor Noise for Replication of a Sound Effects Library," presented at the *129th Convention of the Audio Engineering Society* (2010 Nov.), paper 8218. https://www.eecs.qmul.ac.uk/∼josh/documents/2010/HendryReiss-AES129.pdf.

[99] L. Sebal, *Synthesis of Electric Motors Sounds Based on Physical Models*, Master's thesis, Queen Mary University of London, London, UK (2016 Aug.). http://www.atiam.ircam.fr/Archives/Stages1516/SEBAL_L%C3%A9o_Rapport_Stage.pdf.

[100] Y. Fu, D. Murphy, and A. Southern, "Traffic Flow Auralisation Based on Single Vehicle Pass-by Noise Synthesis," in *Proceedings of the 23rd International Congress on Acoustics*, pp. 1705–1712 (Aachen, Germany) (2019 Sep.). https://pub.dega-akustik.de/ICA2019/data/articles/001031.pdf.

[101] M. Kauth, S. Pischinger, K. Pfleiderer, and M. Nijs, "Engine Noise Synthesis Based on Cylinder Pressure Data," *MTZ Worldwide*, vol. 74, no. 10, pp. 58–64 (2013 Sep.). https://doi.org/10.1007/s38313-013-0103-7.

[102] H. Hacıhabiboğlu, "Procedural Synthesis of Gunshot Sounds Based on Physically Motivated Models," in O. Korn and N. Lee (Eds.), *Game Dynamics: Best Practices in Procedural and Dynamic Game Content Generation*, pp. 47–69 (Springer, Cham, Switzerland, 2017). https://doi.org/10.1007/978-3-319-53088-8_4.

[103] C. Verron, *Synthese Immersive de Sons d'Environnement*, Ph.D. thesis, Aix-Marseille University, Marseille, France (2010 Jun.). https://kronland.

fr/publications/synthese-immersive-de-sons-denvironnement/.

[104] D. Marelli, M. Aramaki, R. Kronland-Martinet, and C. Verron, "Time–Frequency Synthesis of Noisy Sounds With Narrow Spectral Components," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 8, pp. 1929–1940 (2010 Apr.). https://doi.org/10.1109/TASL.2010.2040532.

[105] D. Marelli, M. Aramaki, R. Kronland-Martinet, and C. Verron, "An Efficient Time–Frequency Method for Synthesizing Noisy Sounds With Short Transients and Narrow Spectral Components," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 4, pp. 1400–1408 (2011 Nov.). https://doi.org/10.1109/TASL.2011.2176334.

[106] C. Verron and G. Drettakis, "Procedural Audio Modeling for Particle-Based Environmental Effects," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), paper 8764. https://hal.science/hal-00759818/document.

[107] C. Drioli and D. Rocchesso, "Acoustic Rendering of Particle-Based Simulation of Liquids in Motion," *J. Multimodal User Interfaces*, vol. 5, no. 3, pp. 187–195 (2011 Jul.). https://doi.org/10.1007/s12193-011-0063-7.

[108] S. J. Miklavcic, A. Zita, and P. Arvidsson, "Computational Real-Time Sound Synthesis of Rain," in *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx)*, pp. 169–172 (Naples, Italy) (2004 Oct.). https://huelights.com/docs/P_169.pdf.

[109] S. Liu, H. Cheng, and Y. Tong, "Physically-Based Statistical Simulation of Rain Sound," *ACM Trans. Graph.*, vol. 38, no. 4, paper 123 (2019 Jul.). https://doi.org/10.1145/3306346.3323045.

[110] P. R. Cook, "Din of an "Iniquity": Analysis and Synthesis of Environmental Sounds," in *Proceedings of the 13th International Conference on Auditory Display (ICAD)*, pp. 167–172 (Montreal, Canada) (2007 Jun.). https://hdl.handle.net/1853/49988.

[111] J. Craig, "Adaptive Audio Engine for EEG-Based Horror Game," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), paper 9802.

[112] T. Smyth and J. O. Smith, "Applications of Bioacoustics in Physical Modeling and the Creation of New Musical Instruments," in *Proceedings of the International Symposium on Musical Acoustics (ISMA)*, pp. 267–270 (Perugia, Italy) (2001 Sep.). https://ccrma.stanford.edu/∼tamara/publications/isma2001.pdf.

[113] C. Strandberg, *Mediating Interactions in Games Using Procedurally Implemented Modal Synthesis: Do Players Prefer and Choose Objects With Interactive Synthetic Sounds Over Objects With Traditional Sample Based Sounds?*, Master's thesis, Luleå Tekniska Universitet, Luleå, Sweden (2018 Mar.). https://ltu.diva-portal.org/smash/get/diva2:1191665/FULLTEXT02.pdf.

[114] M. Rath, F. Avanzini, N. Bernardini, et al., "An Introductory Catalog of Computer-Synthesized Contact Sounds, in Real-Time," in *Proceedings of the 14th Colloquium of Musical Informatics*, pp. 103–108 (Firenze, Italy) (2003 May). https://avanzini.di.unimi.it/downloads/publications/rath_cim03_finalsub.pdf.

[115] J. Atherton and G. Wang, "Chunity: Integrated Audiovisual Programming in Unity," in *Proceedings of the NIME*, pp. 102–107 (Blacksburg, VA) (2018 Jun.). https://www.nime.org/proceedings/2018/nime2018_paper0024.pdf.

[116] D. Menzies, "Physically Motivated Environmental Sound Synthesis for Virtual Worlds," *EURASIP J. Audio Speech Music Process.*, vol. 2010, paper 11137878 (2011 Jan.).

[117] M. Rath, "An Expressive Real-Time Sound Model of Rolling," in *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx)*, paper 40 (London, UK) (2003 Sep.).

[118] G. Durr, L. Peixoto, M. Souza, R. Tanoue, and J. D. Reiss, "Implementation and Evaluation of Dynamic Level of Audio Detail," in *Proceedings of the 56th AES International Conference: Audio for Games* (2015 Feb.), paper 3-3. https://www.eecs.qmul.ac.uk/∼josh/documents/2015/Durr%20et%20al%20-%20AES56%20-%202015.pdf.

[119] J. N. Chadwick and D. L. James, "Animating Fire With Sound," *ACM Trans. Graph.*, vol. 30, no. 4, paper 84 (2011 Jul.). https://doi.org/10.1145/2010324.1964979.

[120] H. E. Tez, R. Selfridge, and J. Reiss, "Ocean Wave Sound Synthesis and Perceptual Evaluation," presented at the *151st Convention of the Audio Engineering Society* (2022 May), paper 10565. https://www.eecs.qmul.ac.uk/∼josh/documents/2022/Tez%20AES152.pdf.

[121] M. Rau, O. Das, and E. K. Canfield-Dafilou, "Improved Carillon Synthesis," in *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx)*, pp. 1–8 (Birmingham, UK) (2019 Sep.).

[122] A. Barahona and S. Pauletto, "Perceptual Evaluation of Modal Synthesis for Impact-Based Sounds," in *Proceedings of the 16th Sound and Music Computing Conference*, pp. 34–38 (Malaga, Spain) (2019 May). https://www.smc2019.uma.es/articles/P1/P1_05_SMC2019_paper.pdf.

[123] E. Murphy, M. Lagrange, G. Scavone, P. Depalle, and C. Guastavino, "Perceptual Evaluation of a Real-Time Synthesis Technique for Rolling Sounds," in *Proceedings of the Conference on Enactive Interfaces*, pp. 154–159 (Pisa, Italy) (2008 Nov.).

[124] L. Mengual, D. Moffat, and J. D. Reiss, "Modal Synthesis of Weapon Sounds," in *Proceedings of the 61st AES International Conference: Audio for Games* (2016 Feb.), paper 4-1. https://qmro.qmul.ac.uk/xmlui/bitstream/handle/123456789/13348/Moffat%20MODAL%20SYNTHESIS%20OF%20WEAPON%202016%20Accepted.pdf?sequence=2.

[125] V. Guggiana, A. Darvishi, E. Munteanu, et al., "Analysis and Synthesis of Environmental Sounds," Swiss Federal Institute of Technology Tech. Rep., vol. 95.12 (1995 Jan.). https://pure.tue.nl/ws/portalfiles/portal/4278660/587153.pdf.

[126] M. Athineos and D. P. Ellis, "Sound Texture Modelling With Linear Prediction in Both Time and Fre-

quency Domains," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, paper V-648 (Hong Kong, China) (2003 Apr.). https://doi.org/10.1109/ICASSP.2003.1200054.

[127] J. Zúñiga and J. D. Reiss, "Realistic Procedural Sound Synthesis of Bird Song Using Particle Swarm Optimization," presented at the *147th Convention of the Audio Engineering Society* (2019 Oct.), e-Brief 555. https://www.eecs.qmul.ac.uk/∼josh/documents/2019/20578.pdf.

[128] J. Fornari, "An Evolutionary Algorithm to Create Artificial Soundscapes of Birdsongs," *Int. J. Arts Technol.*, vol. 9, no. 1, pp. 39–58 (2016 Mar.). https://doi.org/10.1504/IJART.2016.075409.

[129] B. O'Brien, B. Juhas, M. Bieńkiewicz, et al., "Considerations for Developing Sound in Golf Putting Experiments," in *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*, pp. 338–358 (Matosinhos, Portugal) (2017 Sep.). https://doi.org/10.1007/978-3-030-01692-0_23.

[130] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone, "A 3-D Immersive Synthesizer for Environmental Sounds," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 6, pp. 1550–1561 (2010 Aug.). https://doi.org/10.1109/TASL.2009.2037402.

[131] C. Heinrichs, A. McPherson, and A. Farnell, "Human Performance of Computational Sound Models for Immersive Environments," *New Soundtr.*, vol. 4, no. 2, pp. 139–155 (2014 Aug.). https://doi.org/10.3366/sound.2014.0059.

[132] L. J. Paul, "The Procedural Sound Design of Sim Cell," presented at the *137th Convention of the Audio Engineering Society* (2014 Oct.), paper 9209.

[133] N. Bonneel, G. Drettakis, N. Tsingos, I. Viaud-Delmon, and D. James, "Fast Modal Sounds With Scalable Frequency-Domain Synthesis," in *Proceedings of the ACM SIGGRAPH*, paper 24 (Los Angeles, CA) (2008 Aug.). https://doi.org/10.1145/1399504.1360623.

[134] J. Quintas, *Procedural Audio Generation Applied to Digital Games*, Master's thesis, Polytechnic Institute of Cávado and Ave (IPCA), Barcelos, Portugal (2019 Mar.). https://winlib.biblioteca.ipca.pt/Nyron/Library/Catalog/winlibsrch.aspx?skey=59E30D421D1B49D0BFB33C2C9D823484&cap=&pesq=5&thes1=22814&doc=11184.

[135] D. A. Jaffe, "Ten Criteria for Evaluating Synthesis Techniques," *Comput. Music J.*, vol. 19, no. 1, pp. 76–87 (1995 Spring). https://doi.org/10.2307/3681301.

[136] T. Tolonen, V. Välimäki, and M. Karjalainen, "Evaluation of Modern Sound Synthesis Methods," Helsinki University of Technology Tech. Rep. 48 (1998 Mar.). http://legacy.spa.aalto.fi/publications/reports/sound_synth_report.pdf.

[137] D. Moffat and J. D. Reiss, "Perceptual Evaluation of Synthesized Sound Effects," *ACM Trans. Appl. Percept.*, vol. 15, no. 2, paper 13 (2018 Mar.). https://doi.org/10.1145/3165287.

[138] A. Horner and S. Wun, "Evaluation of Iterative Matching for Scalable Wavetable Synthesis," in *Proceedings of the AES 29th International Conference: Audio for Mobile and Handheld Devices* (2006 Sep.), paper 7-1.

[139] S. Heise, M. Hlatky, and J. Loviscach, "Automatic Cloning of Recorded Sounds by Software Synthesizers," presented at the *127th Convention of the Audio Engineering Society* (2009 Oct.), paper 7858.

[140] B. Hamadicharef and E. Ifeachor, "Perceptual Modeling of Piano Tones," presented at the *119th Convention of the Audio Engineering Society* (2005 Oct.), paper 6525.

[141] B. Hamadicharef and E. Ifeachor, "Objective Prediction of Sound Synthesis Quality," presented at the *115th Convention of the Audio Engineering Society* (2003 Oct.), paper 5958.

[142] A. Merer, S. Ystad, R. Kronland-Martinet, and M. Aramaki, "Abstract Sounds and Their Applications in Audio and Perception Research," in *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, pp. 176–187 (Málaga, Spain) (2010 Jun.). https://doi.org/10.1007/978-3-642-23126-1_12.

[143] M. Aramaki, R. Kronland-Martinet, and S. Ystad, "Perceptual Control of Environmental Sound Synthesis," in *Proceedings of the International Symposium on Computer Music Modeling and Retrieval (CMMR) and Conference on Frontiers of Research in Speech and Music (FRSM)*, pp. 172–186 (Bhubaneswar, India) (2011 Mar.). https://doi.org/10.1007/978-3-642-31980-8_13.

[144] L. Gabrielli, S. Squartini, and V. Välimäki, "A Subjective Validation Method for Musical Instrument Emulation," presented at the *131st Convention of the Audio Engineering Society* (2011 Oct.), paper 8561.

[145] M. D. Hoffman and P. R. Cook, "Feature-Based Synthesis: A Tool for Evaluating, Designing, and Interacting With Music IR Systems," in *Proceedings of the 7th International Conference on Music Information (ISMIR)*, pp. 361–362 (Victoria, Canada) (2006 Oct.). https://soundlab.cs.princeton.edu/publications/2006_ismir_fbs.pdf.

[146] J. H. McDermott, E. P. Simoncelli, "Sound Texture Perception via Statistics of the Auditory Periphery: Evidence From Sound Synthesis," *Neuron*, vol. 71, no. 5, pp. 926–940 (2011 Sep.). https://doi.org/10.1016/j.neuron.2011.06.032.

[147] J. A. Ballas, "Common Factors in the Identification of an Assortment of Brief Everyday Sounds," *J. Exp. Psychol.: Hum. Percept. Perform.*, vol. 19, no. 2, pp. 250–267 (1993 Feb.). https://psycnet.apa.org/doi/10.1037/0096-1523.19.2.250.

[148] T. L. Bonebright, N. E. Miner, T. E. Goldsmith, and T. P. Caudell, "Data Collection and Analysis Techniques for Evaluating the Perceptual Qualities of Auditory Stimuli," *ACM Trans. Appl. Percept.*, vol. 2, no. 4, pp. 505–516 (2005 Oct.). https://doi.org/10.1145/1101530.1101550.

[149] H. Hahn, *Expressive Sampling Synthesis. Learning Extended Source-Filter Models From Instrument Sound Databases for Expressive Sample Manipulations*, Ph.D. thesis, Université Pierre et Marie Curie, Paris, France (2015 Sep.). https://hal.science/tel-01263656/file/hahn_phd_thesis.pdf.

[150] G. Scavone, S. Lakatos, P. Cook, and C. Harbke, “Perceptual Spaces for Sound Effects Obtained With an Interactive Similarity Rating Program,” in *Proceedings of International Symposium on Musical Acoustics* (Perugia, Italy) (2001 Sep.). https://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=558251D20883F4741DF41D513E0FFB8A?doi=10.1.1.15.1400&rep=rep1&type=pdf.

[151] S. Lakatos, S. McAdams, and R. Caussé, “The Representation of Auditory Source Characteristics: Simple Geometric Form,” *Percept. Psychophys.*, vol. 59, no. 8, pp. 1180–1190 (1997 Dec.). https://doi.org/10.3758/BF03214206.

[152] X. Ma, C. Fellbaum, and P. R. Cook, “SoundNet: Investigating a Language Composed of Environmental Sounds,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1945–1954 (Atlanta, GA) (2010 Apr.). https://doi.org/10.1145/1753326.1753620.

[153] S. Delle Monache, D. Rocchesso, F. Bevilacqua, et al., “Embodied Sound Design,” *Int. J. Hum.-Comput. Studies*, vol. 118, pp. 47–59 (2018 Oct.). https://doi.org/10.1016/j.ijhcs.2018.05.007.

[154] D. Schwarz, A. Roebel, H. Yeh, and A. La Burthe, “Concatenative Sound Texture Synthesis Methods and Evaluation,” in *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx)* (Brno, Czech Republic) (2016 Sep.). https://hal.science/hal-01427393/document.

[155] M. Comunità, H. Phan, and J. D. Reiss, “Neural Synthesis of Footsteps Sound Effects With Generative Adversarial Networks,” presented at the *152nd Convention of the Audio Engineering Society* (2022 May), paper 10583.

[156] J. D. Reiss, H. T. Tez, and R. Selfridge, “A Comparative Perceptual Evaluation of Thunder Synthesis Techniques,” presented at the *150th Convention of the Audio Engineering Society* (2021 May), e-Brief 640.

[157] E. Fineberg, J. Walters, and J. Reiss, “Advances in Thunder Sound Synthesis,” presented at the *152nd Convention of the Audio Engineering Society* (2022 May), paper 10574. https://arxiv.org/pdf/2204.08026.pdf.

[158] C. Heinrichs and A. McPherson, “Mapping and Interaction Strategies for Performing Environmental Sound,” in *Proceedings of the IEEE VR Workshop: Sonic Interaction in Virtual Environments (SIVE)*, pp. 25–30 (Minneapolis, MN) (2014 Mar.). https://doi.org/10.1109/SIVE.2014.7006286.

[159] A. P. McPherson, R. H. Jack, G. Moro, et al., “Action-Sound Latency: Are Our Tools Fast Enough?” in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pp. 20–25 (Brisbane, Australia) (2016 Jul.). https://www.nime.org/proceedings/2016/nime2016_paper0005.pdf.

[160] B. L. Giordano, P. Susini, and R. Bresin, “Perceptual Evaluation of Sound-Producing Objects,” in K. Franinović and S. Serafin (Eds.), *Sonic Interaction Design*, pp. 151–197 (2013 Dec.). https://doi.org/10.7551/mitpress/8555.003.0007.

[161] N. Bottcher, H. P. Martinez, and S. Serafin, “Procedural Audio in Computer Games Using Motion Controllers: An Evaluation on the Effect and Perception,” *Int. J. Comput. Games Technol.*, vol. 2013, paper 6 (MIT Press, Cambridge, MA, 2013). https://doi.org/10.1155/2013/371374.

[162] A. Merer, M. Aramaki, S. Ystad, and R. Kronland-Martinet, “Perceptual Characterization of Motion Evoked by Sounds for Synthesis Control Purposes,” *ACM Trans. Appl. Percept.*, vol. 10, no. 1, pp. 1–24 (2013 Mar.). https://doi.org/10.1145/2422105.2422106.

[163] A. Barahona-Rıos and S. Pauletto, “Synthesising Knocking Sound Effects Using Conditional WaveGAN,” in *Proceedings of the 17th Sound and Music Computing Conference*, pp. 450–456 (Torino, Italy) (2020 Jun.). https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1446839&dswid=-5295.

[164] P. Esling, N. Masuda, A. Bardet, R. Despres, and A. Chemla-Romeu-Santos, “Flow Synthesizer: Universal Audio Synthesizer Control With Normalizing Flows,” *Appl. Sci.*, vol. 10, no. 1, paper 302 (2019 Dec.). https://doi.org/10.3390/app10010302.

[165] A. Vinay and A. Lerch, “Evaluating Generative Audio Systems and Their Metrics,” *arXiv preprint arXiv:2209.00130* (2022 Aug.). https://doi.org/10.48550/arXiv.2209.00130.

[166] H. Caracalla and A. Roebel, “Sound Texture Synthesis Using RI Spectrograms,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 416–420 (Barcelona, Spain) (2020 May). https://doi.org/10.1109/ICASSP40776.2020.9053376.

[167] C. Donahue, J. McAuley, and M. Puckette, “Adversarial Audio Synthesis,” in *Proceedings of the International Conference on Learning Representations (ICLR)*, paper 796 (New Orleans, LA) (2019 May). https://openreview.net/pdf?id=ByMVTsR5KQ.

[168] A. Caillon and P. Esling, “Streamable Neural Audio Synthesis With Non-Causal Convolutions,” *arXiv preprint arXiv:2204.07064* (2022 Apr.). https://doi.org/10.48550/arXiv.2204.07064.

[169] A. Caillon and P. Esling, “RAVE: A Variational Autoencoder for Fast and High-Quality Neural Audio Synthesis,” *arXiv preprint arXiv:2111.05011* (2021 Nov.). https://doi.org/10.48550/arXiv.2111.05011.

[170] N. Devis, N. Demerlé, S. Nabi, D. Genova, and P. Esling, “Continuous Descriptor-Based Control for Deep Audio Synthesis,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5 (Rhodes Island, Greece) (2023 Jun.). https://doi.org/10.1109/ICASSP49357.2023.10096670.

[171] F. Caspe, A. McPherson, and M. Sandler, “DDX7: Differentiable FM Synthesis of Musical Instrument Sounds,” *arXiv preprint arXiv:2208.06169* (2022 Dec.). https://archives.ismir.net/ismir2022/paper/000073.pdf.

[172] N. Masuda and D. Saito, “Improving Semi-Supervised Differentiable Synthesizer Sound Matching for Practical Applications,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 863–875 (2023 Jan.). https://doi.org/10.1109/TASLP.2023.3237161.

[173] C. Verron, M. Aramaki, A. Gonot, et al., “Event-Driven Interactive Solid Sound Synthesis,” in *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research* (Marseille, France) (2013 Oct.). https://www.noisemakers.fr/charlesverron.com/content/papers/2013-interactiveSolids_cmmr.pdf.

[174] D. Menzies, “Phya and Vfoley, Physically Motivated Audio for Virtual Environments,” in *Proceedings of the 35th AES International Conference: Audio for Games* (2009 Feb.), paper 17. https://eprints.soton.ac.uk/371717/1/10.1.1.149.9848.pdf.

[175] K. Cascone, D. T. Petkevich, G. P. Scandalis, et al., “Apparatus and Methods for Synthesis of Internal Combustion Engine Vehicle Sounds,” US Patent 6,959,094 (2005 Oct.). https://patents.google.com/patent/US6959094B1/en.

THE AUTHORS

Dimitris Menexopoulos

Pedro D. Pestana

Joshua D. Reiss

Dimitris Menexopoulos is a versatile composer, sound designer, audio technologist and multi-instrumentalist from Thessaloniki, Greece. He has two solo albums under his name (*Perpetuum Mobile*, 2017; *Phenomena*, 2014), two EPs (*Modern Catwalk Music*, 2022; *40 EP*, 2020), two published soundtracks (*Iolas Wonderland*, 2021; *The Village*, 2019), and various performances internationally. As a designer, he has presented work at sites including the Barbican Centre (Nesta FutureFest, 2019, with Akvile Terminaite), the Somerset House (24 Hours in Uchronia With Helga Schmid, 2020) and Christie’s London (Christie’s Lates, 2023, with Scarlett Yang). His original music creation devices have been demonstrated at venues in the United Kingdom (Iklectik, 2020), France (IRCAM, 2019 and 2020), and the United States (Mass MoCA, 2019). He is currently conducting Ph.D. research on the topic of exploiting game graphics rendering for sound generation at Queen Mary University of London.

•

Pedro Duarte Pestana is currently an Assistant Professor at the Science and Technology Department of the Open University (UAb), after one decade with the Science and Arts department at the Catholic University of Portugal (UCP), where he coordinated the Sound and Image undergraduate degree. He is a researcher at the Center of Science and Technology of the Arts (CITAR) at UCP, of which he was the Director between 2016 and 2018. He also consults as Lead Audio Research at Mixgenius, a company that specializes in intelligent audio production software. He has investigated topics pertaining to perception and cognition in audio, machine learning systems, acoustics, interactive sound design, and digital audio effects. He has been an active member of the Audio Engineering Society for over a decade and won the best paper award at the 134th AES Convention in Rome. He has 17 published papers in journals and conferences with an h-index of 8 and over 250 total citations. He actively collaborates in artistic projects of international visibility and holds one patent.

•

Joshua Reiss is Professor of Audio Engineering with the Centre for Digital Music at Queen Mary University of London. He has published more than 200 scientific papers (including over 50 in premier journals and seven best paper awards) and co-authored three books. His research has been featured in dozens of original articles and interviews on TV, on the radio, and in the press. He is a Fellow and Past President of the Audio Engineering Society and chair of their Publications Policy Committee. He co-founded the highly successful spin-out company, LandR, and recently co-founded Waveshaper AI, RoEx, and Nemisindo, also based on his team’s research. He maintains a popular blog, YouTube channel, and Twitter feed for scientific education and dissemination of research activities.