Human Performance of Computational Sound Models for Immersive Environments

Christian Heinrichs, Andrew McPherson, and Andy Farnell

Abstract

This paper presents a method for incorporating the expressivity of human performance into real-time computational audio generation for games and other immersive environments. In film, Foley artistry is widely recognised to enrich the viewers experience, but the creativity of the Foley artist cannot be easily transferred to interactive environments where sound cannot be recorded in advance. We present new methods for human performers to control computational audio models, using a model of a squeaky door as a case study. We focus on the process of selecting control parameters and on the mapping layer between gesture and sound, referring to results from a separate user evaluation study. By recording high-level control parameters rather than audio samples, performances can be later varied to suit the details of the interactive environment.

Keywords

Computational Audio, Foley, Performable Sound Model, Interaction, Immersive Environments, Evaluation

Note: This paper has been accepted for publication in The New Soundtrack journal (Edinburgh University Press, http://www.euppublishing.com/journal/sound) and is expected to be published in September 2014.

1 Introduction

Synthesised sound offers us a new way of thinking about sound design, and promises a much needed solution to the complex problem of designing sound for dynamic interactive environments. However, the lack of human performance in the currently proposed design processes of computational audio models contradicts what is arguably one of the most integral elements in the art of sound design. Onomatopoeia illustrate the ability to effectively synthesise many sounds using the limited and mechanically unrelated apparatus of the human voice. We re-enact sound in this way not only as a means of referral but as a way of conveying extra layers of meaning associated with the sound or our relationship to the sounding object. Luigi Russolo's 1913 manifesto, *The Art of Noises*, coincided with his building of 16 human performable instruments for producing environmental sounds, despite the availability of recording technology, albeit primitive at the time. Similar performable artefacts were created by Jimmy McDonald to give the sounds in early Disney animations the necessary human component. Even now Foley artistry remains an integral part of most film productions.

Now, as it is becoming possible to render rich and complex environments in real-time without relying on recorded sound, and as sensing technology has also reached a stage where we can extract high-bandwidth signals from our movements and gestures in space, it is important to revisit the notion of human-performed sound.

In this paper we elaborate on the concept of behavioural abstraction proposed by Farnell (2010) and distinguish between *physical models*, *physically-informed models* and our new proposed category of *performable models*. We outline a simple procedure for extending a procedural model with a few physics-based variable parameters, using our new model of a creaking door as a case study. On one hand the extended model is a more flexible asset in that it is capable of producing a wider breadth of door sounds. On the other hand, a larger parameter space freed from physical associations means that the model is more conducive to conventional mapping techniques as used in the realm of digital musical instruments. Finally we describe three mapping strategies between a generic touch-based interface and the newly parametrised model, including a physically-inspired control layer that is based on metaphorical interaction borrowed from a related sound-producing mechanism. Differences between each mapping strategy are outlined alongside results from a user evaluation study.

Ultimately, performable models could play an important role in the implementation of computational audio into interactive scenarios. The use of performed parameter sequences lends itself well to techniques well known from animation (e.g. blending) and offers an attractive compromise between sampled audio, which is too inflexible for interactivity, and purely physical models, where encoding human expressivity into the model's behaviour is problematic.

1.1 Computational Audio

Established sound design techniques based on the recording and treatment of sound remain effective and irreplaceable tools for films, animations and other linear media. However, in games and interactive applications, so-called non-linear media, the exclusive use of sampled sound is widely known to have many shortcomings. While the sound quality of the samples themselves can match what is heard in the cinema, the interactive nature of games leads to issues of repetition and misalignment to generative movements and player controls. Currently these issues are mitigated through the use of large sample libraries and the coarse manipulation of these samples by means of cross-fading, randomization and granular techniques. These measures result in a slightly more varied sound image but are not enough to constitute a truly responsive environment. A new paradigm is promised by the notion of *computational audio*, which in this paper will be used to refer to the digital generation of sounds with minimal or no direct use of recorded samples.

On one hand numerical techniques are capable of accurately simulating increasingly complex sounds in real-time (Zheng & James 2011, Serafin 2004) and exhaustive methods (e.g. finite-difference time-domain) are gradually getting closer to real-time implementations (Bilbao et al 2013). This approach to synthesis (more generally referred to as *physical modelling*) is particularly useful when sound needs to be tightly coupled to a moving image or physical action, where compatible physical parameters for driving the model are available. For example, the *Sound Design Toolkit* (Delle Monache et al 2008a) takes Gaver's framework of ecological acoustics (Gaver 1993) as a fundamental principle in coupling everyday gestures to simulated sounds corresponding to actions such as 'crumpling' or 'rolling'.

On the other hand the increased understanding and teaching of tools such as PureData[1] and Max/MSP[2] make it possible to design sounds using smaller sound-generation components (such as oscillators and filters), giving the designer greater artistic freedom over the resulting sound. Here, the design is often aided by close observation of physical processes underlying sound sources as they occur in the real world, however does not need to pertain to physical laws in a mathematically precise way. Aside from the aesthetic benefits this approach also makes it possible to design simplified models, which can be crucial in a situation where computational resources are scarce. This approach is referred to as *physically-inspired* *modelling* and comprehensive overviews of techniques and implementations can be found in textbooks by Cook (2002) and Farnell (2010).

As with physical models, a set of higher-level parameters can be used in order to map actions to sound events produced by a physically-inspired model. *Behavioural abstraction* as introduced by Farnell (2010) refers to an abstract high-level component that acts as an interface between an interactive application and internal parameters of such a model. In most cases the designer will decide on these specific behavioural dimensions before implementing the model. This can be very effective in situations where realism or consistency (e.g. mechanical) is an important factor and when the listener needs to extract quantifiable information from the sound (e.g. contact sounds).

In other situations where the sound contradicts the moving image or is a heightened representation – in the sense of Chion's audio-visual contract (Chion, 1994) – 'hard-coding' the behaviour into the model may be insufficient to produce the desired effect. In the conventional workflow these sounds are often handled by a Foley artist who performs the desired sound using objects that aren't necessarily related to the corresponding source but are instead perceptually relevant to the listener. These sounds are usually synchronised directly to the image by the performer in order to capture subtle expressive nuances.

The resulting behaviour is hard to capture computationally. The difficulty lies in preempting the expressive nuances that arise from performance in the design process. A useful example is the sound of footsteps. It is a common misconception that larger persons will produce a heavier sounding footfall when in reality this is very often not the case. Nonetheless such misrepresentations will often be perceived as more believable and so using the direct sound from the filmed actors can produce a bland and disjointed sound image (Ament 2012). The same problem applies to the use of physical parameters derived from a game engine (i.e. body mass, shoe type, surface properties, and so on) to drive a sound model. While the Foley artist can easily create a more believable soundtrack by simply performing the footsteps in synchrony to the moving image, this kind of immediacy is currently lacking in the design of computational models.

Performative strategies for models of naturally occurring sounds have been the subject of some research in the context of both Digital Musical Instruments and Sonic Interaction Design. Essl and O'Modrhain (2006) propose an 'enactive' approach to designing musical instruments, where interactions with everyday objects are captured using sensing technology and used to drive alternative sound models. For example, in one of their implementations a whiteboard eraser is used to perform friction sounds produced by physically-inspired models (Essl and O'Modrhain 2005). This is achieved by driving behavioural parameters of the model using sensor data collected from contact microphones and a pressure sensor. A similar approach is taken by Delle Monache et al. in their implementations of *Sonically Augmented Found Objects* (Delle Monache et al 2008b), where interactions with everyday objects (such as cutlery and jugs) are augmented using contradictory sounds.

This is an elegant solution to the control of sound models, due to the tight coupling between natural everyday gestures and sound events. While this approach produces a very natural sense of control over the sound it relies on the use of sound models with a pre-defined behaviour. In other words, in order to implement a natural tangible interaction, gestures need to be coupled with physical parameters corresponding to ecologically meaningful sonic behaviours such as 'crumpling', 'scraping', and so on. In contrast, we are interested in using performance as a means of *defining* these behaviours and therefore this is likely to entail working with a more abstract set of parameters. To borrow Pierre Schaeffer's term these are cases where the sound designer might take on a *reduced listening* approach (Chion, 1983) to control particular aspects of the way the sound evolves over time while ignoring connotations that exist outside the pure essence of the sound (e.g. relating to physics or gesture).

1.2 The Behavioural Layer

All of the above-mentioned manifestations of computational audio share the property of being controlled by a pre-defined set of high-level parameters. Depending on the application, these can correspond to:

- A person’s physical movement in space
- Interactions between virtual objects, a virtual environment and a user
- Game states
- Abstract datasets

In most cases there will be two types of parameters. The first type is a set of constants that corresponds to the physical properties of the object being modelled, including size, material, and so on. These are often referred to as *fixed* or *invariant* parameters (Gaver 1993, Casey 1998). The second type is a set of variables – or *dynamic* parameters – whose values change over time, such as velocity, applied force and position. In the case of a physical model these parameters form part of a complex algorithm (usually a partial differential equation) that is solved for every new digital sample of audio that is generated (see Figure 1a). Here, the design is guided by mathematical principles, with its output often not being auditioned before the algorithm has been fully implemented. In contrast, a physically-inspired model is typically designed using smaller components with known acoustical behaviours, which enables sound to be a guiding principle in the process, as much as an understanding of the underlying physics. This results in an inherent signal processing chain that is capable of producing a wide range of sounds including those associated with the physical behaviour being modelled. The parameters that drive individual elements of this signal chain relate to acoustic effects such as the frequencies of oscillators, the resonance of filters, gain levels, and so on. In order to narrow the model down to a specific set of behaviours there needs to be an intermediary parametrisation layer that transforms the top-level parameters (i.e. fixed and dynamic parameters) into lower-level ones that control the signal chain. We refer to this as the *behavioural layer* of the model, because the desired behaviour is encoded into the parametrisation of the signal-processing chain.

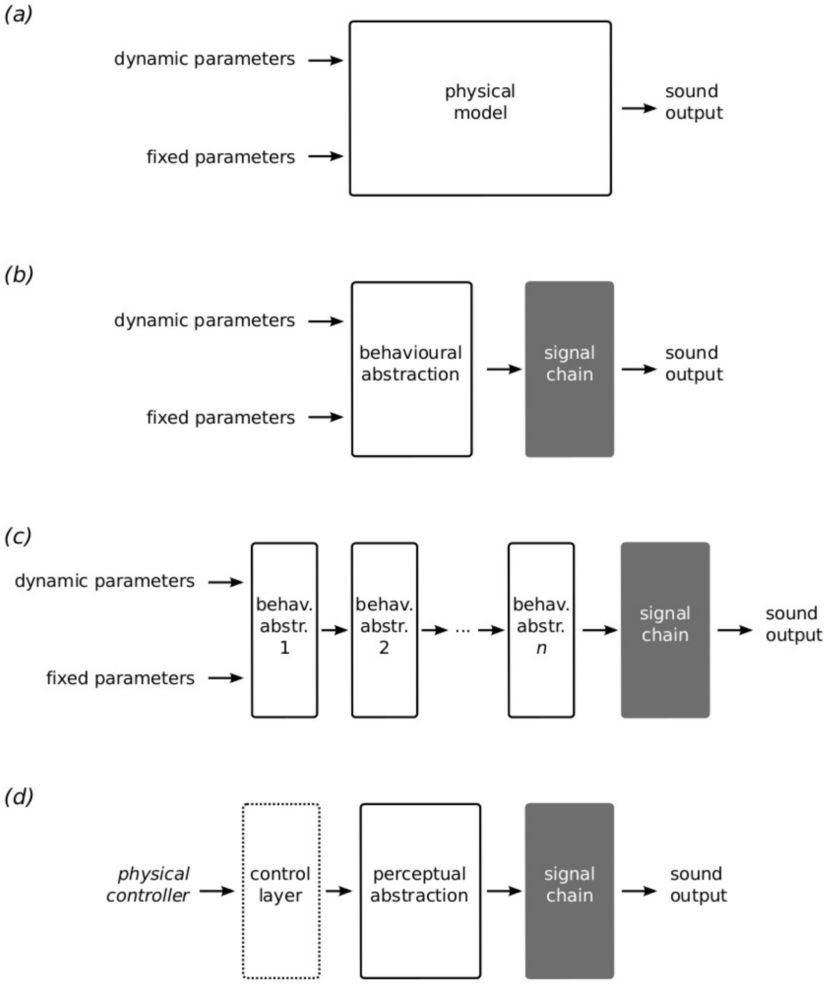

Figure 1: Overview of layered Sound Model Structures: (a) Intermediate Layer for a Physical Model, (b) Behavioural Abstraction in a Physically-Informed Model, (c) Multiple Layers of Abstraction in a Physically-Informed Model, (d) Intermediary Layers for proposed Performable Model (referred to in Section 3).

1.3 Complexity of the Behavioural Layer

Depending on the choice of top-level parameters the model's behaviour might be best represented by several abstraction layers chained together (see Figure c). Taking the case of a creaking door for example, instead of supplying a continuous force parameter the targeted behaviour might be for the door to spring open at a specified intensity. In this case we would create a new behavioural layer that generates changes in applied force over time based on the value of a top-level *intensity* parameter. Narrowly specified behaviours require more complex abstraction layers and result in a more constrained sound output. On the other hand a less complex or missing abstraction layer results in a much larger range of potential behaviours expressible by the model, but at the cost of a larger and more abstract parameter space.

In many ways physically-informed models offer a lot of creative potential to the sound designer through the ability to 'tap in' and extend or reduce the model to suit their needs. This would be hard to achieve with a physical model which is constrained from the outset to the specific behaviour it was designed to simulate. On the other hand there are still some creative drawbacks to this approach. Extension through addition of behavioural layers allows more specificity and useful parameter reduction but reduces the flexibility and therefore the interactivity of the sound. This brings us back to issues of repetition and unresponsiveness, as in pre-recorded samples. Furthermore, designing unique and expressive behaviours is difficult as the design process lacks the immediacy available to Foley artists in the traditional sound design workflow. Extending the model to spring open at different intensities is easy from a physical-analytical point of view, but what if we want it to spring open in *that particular way*? This would require a different set of guiding principles, focused on creating an abstraction layer that enables the design of expressive behaviours through the immediacy of human performance.

In the next sections we will outline a case study of such an implementation: a physically-informed model of a squeaky door that has been extended to be performable by a human. We outline the design process of developing performable models and how the process of abstraction is likely to differ from the normal approach of encoding physically informed behaviour into the parametrisation of a computational model. We then describe factors to consider in implementing physical control and methods of evaluating performable models, drawn from the field of digital musical instruments (DMIs). Finally results from an evaluation study are presented, testing the relative effectiveness of three different control layers based on our proposed evaluation metrics.

2 Physically-Informed Model of a Squeaky Door

The sound of a squeaking door is caused by the dynamic process of stick-slip friction. The motion of an object weighing down on a surface as tangential force is applied is characterized by a series of semi-regular bursts of velocity. These are caused by dynamic interactions between opposing forces (friction and applied force), where the frictional force changes as a function of the object's velocity. Serafin (2004) has outlined a variety of mathematical models that can be used to model this process including a series of effective dynamic models. These dynamic models employ differential equations to approximate the temporal behaviour caused by stick-slip friction at a very fine level of detail.

While this approach produces highly convincing results, it is very difficult to control the sound beyond the physical parametrisation of the model. The dynamic variables of normal force and tangential force lend themselves well to a literal implementation, for example if one were to use an actual door equipped with appropriate sensors to drive the physical parameters of the model. Otherwise, with a physical model, it is harder to control specific features in the evolution of the sound, whereby the sound designer is more likely to think in terms of timbral variations such as 'pitch' and 'roughness'.

Instead it is possible to take a physically-inspired approach and model the resulting velocity of the object instead of the underlying complex behaviour that has caused it. A basic implementation is outlined by Farnell (2010). The regular bursts of velocity resulting from the object's 'slipping' from and 'sticking' against a surface are simulated using an impulse train generator. The output of the impulse generator is then passed through a bank of band pass filters and delay-lines to simulate the effect of the wooden panel. The frequency and amplitude of the impulses relates broadly to the amount of tangential force applied to the object (in this case resulting in the door's rotation around the hinge). Thus an incoming force parameter is scaled to the desired frequency and amplitude ranges. Finally some temporal smoothing is applied to the force parameter to simulate the mass of the door and resulting momentum. These two stages of smoothing and scaling can be thought of as the behavioural abstraction layer of the door creaking model, while the impulse generator, formant bank and resonators constitute the model's signal processing chain (see Figure 2).

The threshold between behavioural abstraction and the signal processing chain is often ambiguous, in which case one needs to make a subjective decision. A useful guiding principle is to identify *signature processes*, in other words, characteristics that are common across all sounds within a given class. For example, periodic impulses can be observed in all manifestations of stick-slip friction sounds and therefore a pulse-train oscillator along with its variable *frequency* parameter can be regarded as a required element of the sound model. The way frequency varies over time depends on the design of the behavioural layer, however the parameter itself will always exist as part of the model's signal chain.

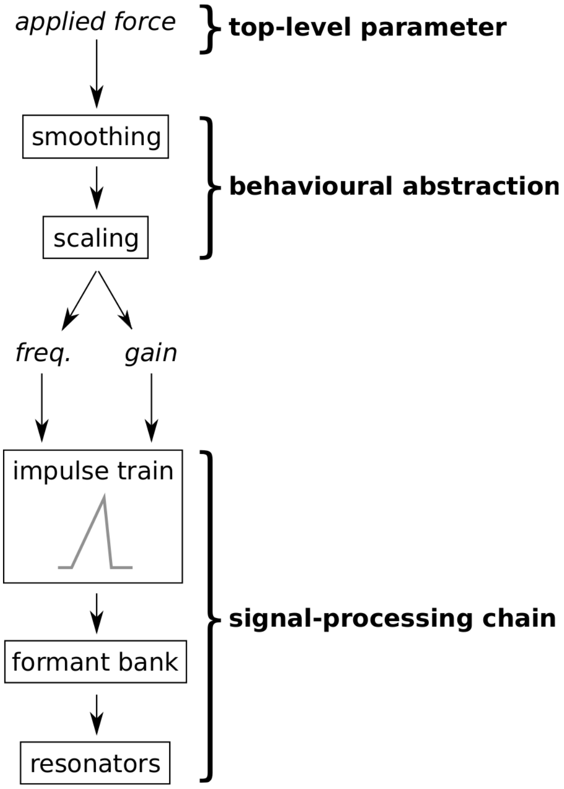

Figure 2: Structure of initial Physically-Informed Door Creaking Model.

2.1 Towards a performable model

When performing the sound of physical sources as part of a design process we are unlikely to be thinking solely in terms of physical behaviours and might instead think along more perceptually relevant dimensions such as ‘brightness’, ‘pitch’, ‘harshness’ and ‘loudness’ (Vicario 2003). Dimensions might be relevant to a wide range of people, or they could be highly subjective and describe a particular feature that is important to the sound designer. We can call such terms *phenomenological* as they refer specifically to the experience of perceiving sound events produced by the object. We therefore propose a different analytical approach in extending the model for performance, namely one where we are listening for differences in the way something *sounds* rather than the way something *works*.

The process starts with the compilation of a library of sounds that together form a suitable sonic description of the model. For our squeaky door model we acquired a large variety of door squeaking and creaking recordings from sample libraries as well as personal recordings. Next, the following steps are carried out iteratively:

1. Choose a sound from the acquired library
2. Attempt to match recording using available model parameters
3. Identify unaccounted perceptual feature
4. Implement and parametrise missing feature
5. Evaluate implementation of feature

Parameter identities ranged from perceptual descriptions of the resulting sound (*roughness*, *brightness*) to more technical descriptions relating to the way they were implemented in the signal chain, for example *amplitude modulation frequency*. Some features required adding new components to the signal chain, such as a noise generator to account for the less regular sound of surface roughness that was found in our analysis to play a prominent role. The dynamic parameters were doubled at one point to account for a second hinge.

Our final model is illustrated in Figure 3. Because our targeted sound library consisted of a broad behavioural range of door squeaks our extended model was correspondingly complex – extending the original single dynamic parameter to seventeen dynamic and more than fifty fixed parameters. In a typical design scenario a computational model is likely to have a more narrowly defined target sound palette – the library being a set of highly stylized concept sounds rather than arbitrary recordings from the everyday environment – resulting in fewer parameters.

The key difference to designing specific behavioural abstractions is that the final parameters are unknown and representative of a subjective interest in sound qualities rather than variations in physical behaviour. Rather than constricting the model the parameter space is expanded until the model is capable of reproducing all the desired sounds, usually resulting in a larger sound output range than before. An advantage of this approach is that, due to the iterative nature of the parametrisation task, parameters are likely to be linearly independent. In other words, the model designer is unlikely to implement a new parameter that can be recreated using a combination of existing parameters. The drawback, of course, is that the sound output range is more likely to contain unwanted sounds as the model's dimensionality increases due to unforeseen parameter combinations. It should also be noted that the parametrisation becomes highly subjective: what one person holds to be an independent perceptual dimension might not apply to the way another person understands or perceives the sound.

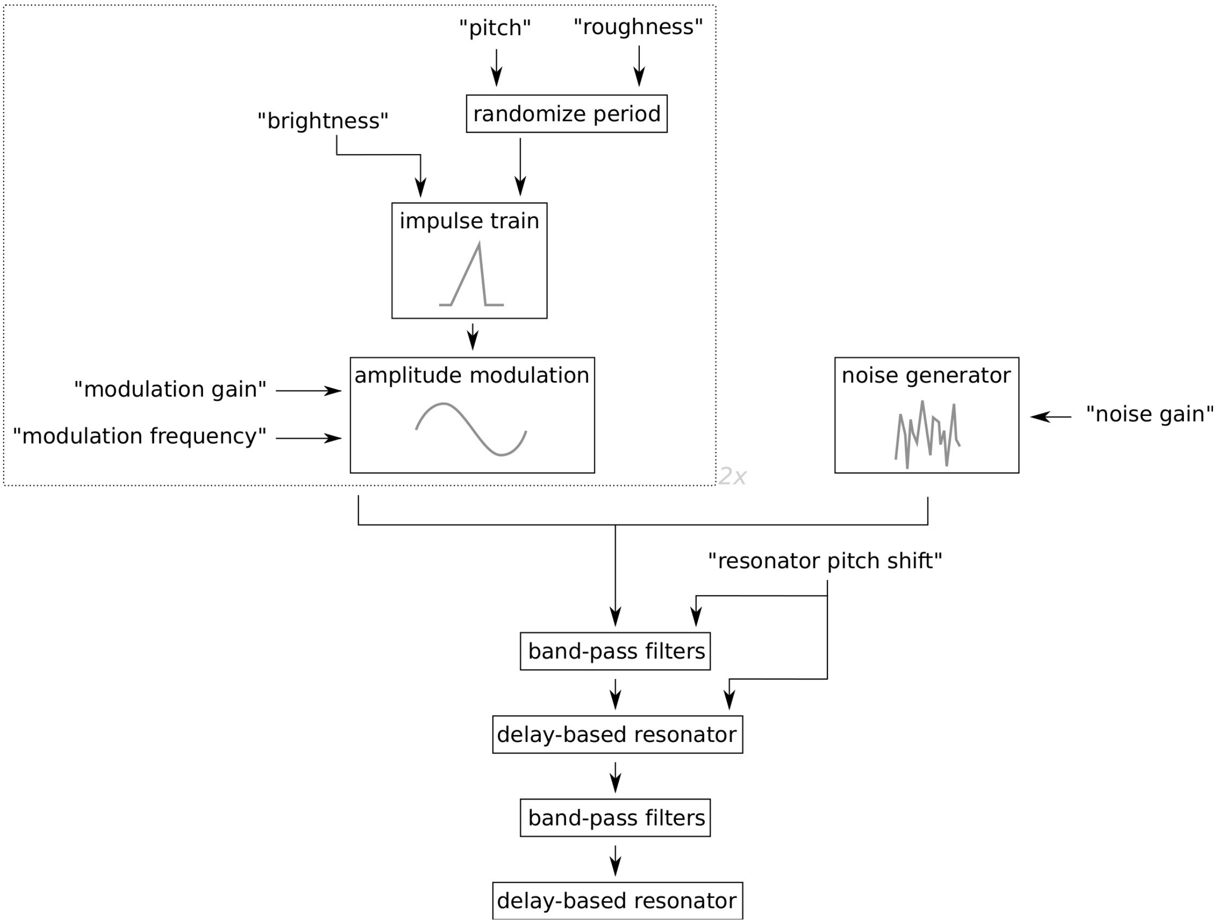

Figure 3: Structure of the extended Squeaky Door Model (variable parameters are indicated in double quotation marks).

3 Designing the Control Layer

At this stage the model has a high-dimensional parameter space that, given the appropriate means of navigation, can produce any imaginable combination of door sounds from our originally targeted sound library. The navigation of the sound model's parameter space is now a problem of dimensional reduction, as we would ideally need to find a means of controlling up to seventeen parameters simultaneously with no more parameters than our brain is capable of processing. Dimensional reduction is a problem faced continually in musical interaction, particularly in the development of digital musical instruments. The success of overcoming this problem depends on an intermediary mapping layer that translates parameters from a physical controller into a larger set of parameters driving a synthesis model. Thus the performable model will consist of two types of intermediary layers: a *control layer* and a *perceptual abstraction layer* (see Figure 1d).

3.1 Mapping Strategies

Hunt and Wanderley (2003) outline four modes of what they term *explicit mappings* between performer actions and synthesis parameters: *one-to-one*, *one-to-many*, *many-to-one* and *many-to-many*. In a one-to-one mapping, each synthesis parameter is controlled by an independent physical input parameter. A useful analogy to this is a mixing desk, where each potentiometer controls a separate gain value. One-to-one mappings become impractical when dealing with large parameter spaces (due to cognitive load), or when there are fewer physical input parameters than synthesis parameters. One-to-many mappings map a single control parameter to several synthesis parameters and many-to-one mappings use more than one control parameter to control a single synthesis parameter. Most acoustic musical instruments would fall into the latter two categories. For example, the pitch of a trombone is controlled by the performer's embouchure as well as the position of the slide and so can be understood as a many-to-one mapping. Variation of energy does not only change the overall volume but also affects the sound quality, implying a one-to-many system.

This terminology is only useful for the most simple of implementations, and – as exemplified in the case of acoustic instruments – can quickly become too ambiguous when dealing with more complex controller-synthesiser couplings. Alternatively, Hunt and Wanderley suggest the use of intermediary abstraction layers that map abstract parameters that are meaningful for a particular controller to a separate set of abstract parameters specific to the synthesis model. Another example of an intermediary mapping is proposed by Henry (2004), who created a collection of objects for the audio programming environment PureData that use mass-spring systems to recreate physical behaviours. These are designed to be used as intermediary layers that treat performed actions as physical actuators of a virtual dynamic system, which in turn produces a natural evolution of synthesis parameters. More recently, machine learning techniques have become an area of interest in designing mapping layers for instruments – for example the Gesture Follower developed by Bevilacqua et al (2009) uses a combination of Hidden Markov Models and Dynamic Time Warping to learn and recognise gestures, allowing variations such as size and speed to control synthesis parameters in realtime.

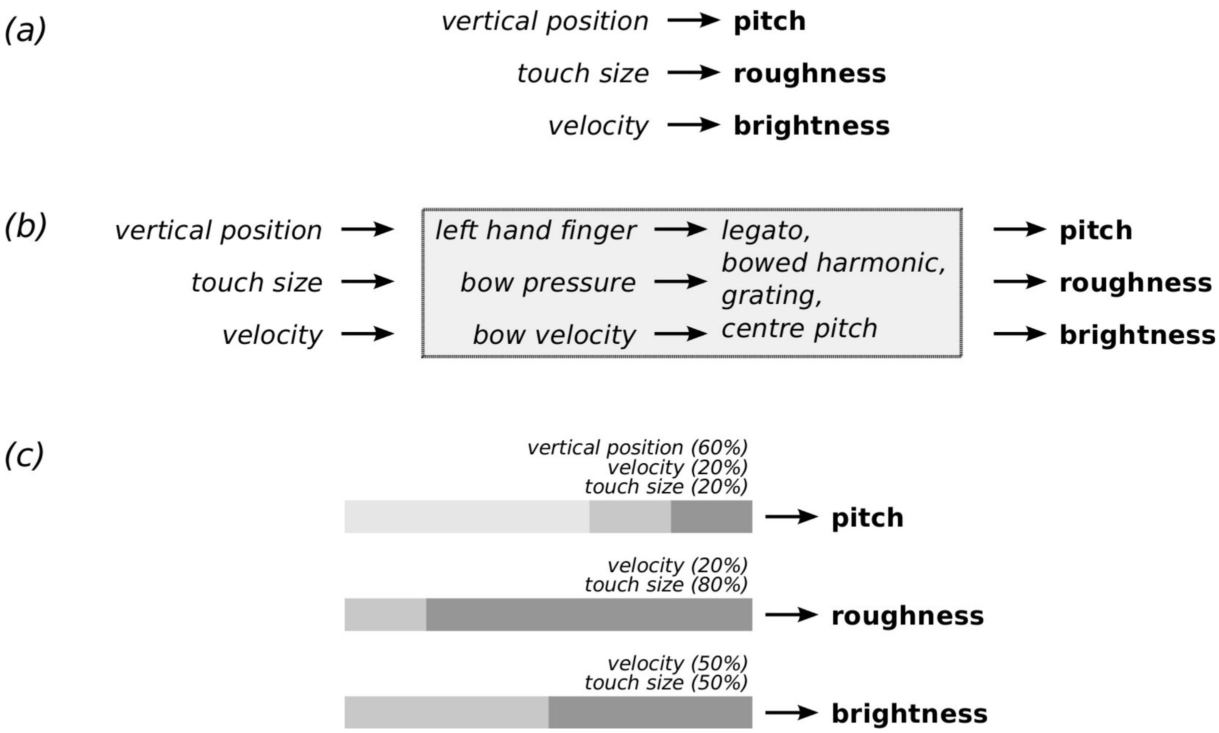

Figure 4: Control Layers for the Squeaky Door Model: (a) One-to-one mapping, (b) physically-inspired Control Layer, (c) Many-to-many mapping

3.2 Control Layers for the Squeaky Door Model

As a preliminary inquiry into the suitability of control layers we restricted the performable interface to three varying synthesis parameters. This allowed us to carry out a more focused evaluation of mapping strategies while keeping the dimensionality of both control and synthesis parameters constant. For this study, we decided to use a trackpad with two axes and touch size to translate human movements into a set of values that vary over time, it being a widely used and a generic style of controller that many people are already familiar with (for example in smartphones, laptops and graphics tablets).

Vertical position, touch area and velocity were used to control pitch, roughness and brightness parameters of the squeaky door model. Three control layers were implemented, each taking a different approach to the way the input dimensions are translated into synthesis parameters.

The first control layer consists of a simple one-to-one mapping, where vertical position controls pitch, touch size controls roughness and velocity controls brightness (see Figure 4a).

The second implementation is what we refer to as a *physically-inspired control layer*. In the same way that it would be possible to reintroduce a physically-informed behavioural layer on top of the new perceptual abstraction of our model we can also implement a physically-inspired layer that is not directly related to the mechanism implied by the model. In this case a control layer based on the interaction between a string and a bow was implemented (see Figure 4b). While both mechanisms involve stick-slip friction, the sounding action of bowing a string is more transparent than the interaction between a door and a hinge. One doesn't need to have played the violin to understand the combined effect that bow pressure and velocity have on the vibration of the string: low velocity and a high amount of pressure results in an unpleasant grating sound, high velocity and low pressure results in a harmonic, and the right combination of the two parameters produces a clean 'legato' tone. Each of these effects can be emulated using the three active synthesis parameters of the squeaky door model: 'grating' is achieved by a combination of low pitch and high roughness (causing the pitch to be unstable) and the upper harmonic can be emulated by doubling the central pitch and lowering the brightness value (reducing the harmonic content of the sound). The central pitch is adjusted by changing the vertical position of the 'bowing' motion.

The final implementation is a many-to-many mapping, in which the sum of multiple control parameters affect each synthesis parameter (see Figure 4c). Weightings of each control parameter are arbitrary and based to some extent on the other two control layers.

3.3 Evaluating Performable Models

We see four features as being particularly important in the design of a performable model: *range*, *nuance*, *repeatability* and *believability*. On one hand the interface should be capable of producing a wide range of sounds, which can easily be explored by the performer. On the other hand the interface should allow a reasonably high degree of control over subtle nuances of the sound, such that when a particular range of sounds has been targeted they can be easily performed and varied. Finally, the interface must be capable of producing sounds that are believable for the given environmental or narrative context.

Nuance and *range* correspond to the degree to which the interface allows to create subtle variations of a single sound and to produce a variety of different sounds within the limits of the model, respectively. These are based on the concepts of *micro-diversity* and *mid-diversity* featured in a comprehensive overview of instrument design issues presented by Jordà (2004). *Repeatability* – originally termed *controllability* by Wanderley and Orio (2002) – is the degree of precision with which a performer can repeat a previously performed sound or sequence of sounds. While the first two metrics correspond to the *variability* of range and nuances, repeatability measures the amount of *control* that the performer has over them.

Finally it would be useful to establish a metric that measures the effectiveness of the interface in the specific context of its implementation. Here we suggest *believability* as a property that is broad enough to apply to a wide range of multimedia environments. Unlike *realism*, this is the extent to which the sound is believed to be part of the environment or an inherent interaction, rather than the degree to which it resembles any physical mechanism or behaviour it may correspond to.

3.4 User Evaluation Study and Results

Each of the three control layers described in Section 3.2 were evaluated in a user study involving fifteen participants with experience in sound design and musical performance. *Range*, *nuance* and *repeatability* metrics for each interface were extracted from subjective surveys and a systematic study that involved performing sound effects for evocative narrative scenarios. *Believability* was tested in a separate listening study featuring a radio play in which sounds varied randomly between performances by previous participants and 'real doors', parameter sequences tuned by hand to match recordings of actual doors. Participants were asked to distinguish between sound effects that they believed to be performed and those believed to be modelled on real doors. We will present the key findings of the study here, referring the reader to (Heinrichs & McPherson 2014) for a detailed overview of the procedure and results.

Participants consistently preferred the many-to-many mapping. Correspondingly, it received the best overall (quantitative and qualitative) ratings for *nuance* and *repeatability* and second-best for *range*. Conversely, the physically-inspired control layer produced the worst ratings for each of these three metrics and was also found to be the least favourite interface and the most challenging to play.

Participants found it hard to distinguish between sound effects thought to be performed by humans and ones thought to be modelled on real doors, however all were thought to be believable when prompted in questionnaires. Interestingly, while the many-to-many mapping was found to be the most effective control layer based on performer-centric metrics, the listening study showed that participants could more easily identify door squeaks performed on this interface than on any others, implying a lower *believability* rating.

4 Implications

The high performer-centric ratings of the control layer based on an arbitrary many-to-many mapping strategy are particularly surprising in comparison to the poor outcomes of the physically-inspired control layer (based on a known behavioural control mechanism) and the one-to-one mapping (providing the most direct relationship between the three parameter pairs). One important aspect (that is beyond the scope of this paper) is that the creative process does not only consist of performing a pre-defined sound effect that exists in the mind of the designer; it also includes an element of exploration which is crucial to the process of imagining the desired sound. A many-to-many mapping that provides lots of access points for the manipulation of perceptual features of the sound may well be more conducive to this process, considering the limited exposure the participants had to each interface. The physically-inspired control layer was also found to be the most challenging to play, suggesting that stronger familiarity with the interface might lead to different results.

On the other hand, the interface that yielded the best results from a performer-centric point of view was more easily recognized as having been performed by a human. This suggests that a control layer that *feels right* to the performer may not always produce results that *sound right* to an external observer.

4.1 Looking ahead: Implementation into Interactive Environments

While implementing performed computational audio is straightforward for linear media such as film, radio and theatre, an added technological layer is required for interactive environments. Each performance results in a sequence of parameters controlling the perceptual layer of the model. On the most basic level, such sequences could be implemented in a similar way that audio samples are treated in games and other interactive media, where samples are triggered based on corresponding events with implied behaviours. On the other hand, parameter sequences also lend themselves to more interesting implementation techniques. For example, interpolating between two sets of sequences will produce more believable results than a simple cross-fade between two audio samples. An analogy can be drawn to parametric animation techniques, where multiple animation sequences are blended together to create a new animation representing an intermediary behaviour. Another approach would be to use a hidden layer of artificial intelligence to learn performed sequences and match them to incoming parameters from the virtual environment. This would result in a virtual Foley artist that mimics the creative decisions of the sound designer, given that enough training data has been provided.

Each of these approaches opens the door to entirely new ways of thinking about and working with sound, integrating the sound designer more tightly into the complex, multi-disciplinary workflow of designing interactive environments.

5 Summary

We have presented a new approach to the design of computational audio models, which aims to incorporate human expressivity through performance. This involves a different way of abstracting the core signal-processing components, focusing on perceptual features rather than varying parameters derived from physical behaviour. As this process is guided by a predefined set of exemplary sounds the result is a model that is capable of producing a wide range of sounds through the manipulation of what are likely to be linearly independent perceptual parameters specific to the model's designer. The requirement of a separate control layer that maps physical movement to this perceptual abstraction layer introduces a new set of design challenges. We have suggested a set of performer-centric metrics for evaluating a control layer, including *range*, *nuance* and *repeatability* which measure the ease of navigating the multidimensional space. A separate metric accounting for the *believability* of the model's sound output is also required. Notably, user studies showed that an interface offering good levels of control over the perceptual space of a model does not necessarily lead to believable results.

The work presented here is just the first step in a greater pursuit of reintroducing human expressivity into the design and implementation of computational audio. It is important to stress that a performable model should not be understood as a 'one-size-fits-all' solution to the performance of a given sound phenomenon. The creation of intermediary layers between a physical controller and the synthesis model needs to be solved creatively, just like all other parts of the design process involved in creating computational models. Nonetheless, systematic inquiries can help design new frameworks and tools that assist the sound designer in this process. Further down the line, the benefits of performable models go beyond *in situ* expressive performance and could potentially introduce novel ways of integrating sound into interactive environments.

Acknowledgments

This work is supported by the EPSRC Doctoral Training Centre in Media and Arts Technology at Queen Mary, University of London.

Notes

1. http://puredata.info/

2. http://cycling74.com/products/max/

Captions

Figure 1 Overview of layered Sound Model Structures: (a) Intermediate Layer for a Physical Model, (b) Behavioural Abstraction in a Physically-Informed Model, (c) Multiple Layers of Abstraction in a Physically-Informed Model, (d) Intermediary Layers for proposed Performable Model (referred to in Section 3).

Figure 2 Structure of initial Physically-Informed Door Creaking Model.

Figure 3 Structure of the extended Squeaky Door Model (variable parameters are indicated in double quotation marks).

Figure 4 Control Layers for the Squeaky Door Model: (a) One-to-one mapping, (b) physically-inspired Control Layer, (c) Many-to-many mapping

References

Ament, V. T. (2009). *The Foley Grail: The Art of Performing Sound for Film, Games, and Animation*. Oxford: Focal Press.

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F. & Rasamimanana, N. (2009). Continuous realtime gesture following and recognition. In *Gesture in embodied communication and human-computer interaction.* Springer, pp. 73–84.

Bilbao, S., Hamilton, B., Torin, A., Webb, C., Graham, P., Gray, A., Kavoussanakis, K. & Perry, J. (2013). Large Scale Physical Modeling Sound Synthesis. In *Proceedings of the Stockholm Musical Acoustics Conference/Sound and Music Computing Conference*. Stockholm, Sweden.

Chion, M. (1983). *Guide des objets sonores: Pierre Schaeffer et la recherche musicale*. Paris: Buchet-Chastel Editions.

Chion, M. (1994). *Audio-Vision: Sound on Screen*. New York: Columbia University Press.

Cook, P. R. (2002). *Real Sound Synthesis for Interactive Applications*. Natick, MA, USA: A. K. Peters, Ltd.

Delle Monache, S., Devallez, D., Drioli, C., Fontana, F., Papetti, S., Polotti, P., & Rocchesso, D. (2008). *Sound synthesis tools for sound design*. Deliverable. Retrieved from http://closed.ircam.fr/uploads/media/closed_deliverable_2.2_public.pdf

Delle Monache, S., Polotti, P., Papetti, S., & Rocchesso, D. (2008). Sonically augmented found objects. In *Proceedings of the 2008 Conference on New Interfaces for Musical Expression,* pp. 154-157. Genova, Italy.

Essl, G., & O'Modhrain, S. (2005). Scrubber: an interface for friction-induced sounds. In *Proceedings of the 2005 Conference on New Interfaces for Musical Expression,* pp. 70–75. Vancouver, Canada.

Essl, Georg, & O'Modhrain, S. (2006). An enactive approach to the design of new tangible musical instruments. *Organised Sound*, *11*(03), pp. 285-296.

Farnell, A. (2010). *Designing Sound*. Cambridge: MIT Press.

Gaver, W. W. (1993). What in the world do we hear? an ecological approach to auditory event perception. *Ecological Psychology*, *5*, pp. 1–29.

Heinrichs, C. & McPherson, A. (2014). Mapping and Interaction Strategies for Performing Environmental Sound. *1st Workshop on Sonic Interactions for Virtual Environments at IEEE VR 2014.* Minneapolis, MN. In press.

Hunt, A., & Wanderley, M. M. (2003). Mapping performer parameters to synthesis engines.

Organised Sound, *7*(02), pp. 97–108.

Jordà, S. (2004). Instruments and Players: Some Thoughts on Digital Lutherie. *Journal of New Music Research*, 33(3), pp. 321–341.

Russolo, L. [1913] (1987). *The Art of Noises*. New York: Pendragon Press.

Serafin, S. (2004). *The sound of friction: real-time models, playability and musical applications*. PhD thesis, Stanford University.

Vicario, G. B. (2001). Prolegomena to the perceptual study of sounds. In Polotti, P., Papetti, S., Rocchesso, D., & Delle, S. (Eds.), *The sounding object (Sob project)* (p. 13). Verona: University of Verona, pp. 17–31.

Wanderley, M. M., & Orio, N. (2002). Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal*, *26*(3), pp. 62–76.

Zheng, C., & James, D. L. (2011). Toward high-quality modal contact sound. In *ACM Transactions on Graphics (TOG), 30*. New York: ACM.

Contributor Details

Christian Heinrichs

Email: *c.heinrichs@qmul.ac.uk*

Christian Heinrichs is a PhD student at the Centre for Digital Music, Queen Mary University of London. His research focuses on computational audio models for games and immersive environments and the reintroduction of human expressivity into the design process of multimedia artefacts. He teaches Sound Design at Queen Mary and has worked on various projects as a sound designer, performer and composer.

Andrew McPherson

Email: *a.mcpherson@qmul.ac.uk*

Andrew McPherson is a Lecturer (Assistant Professor) in the Centre for Digital Music at Queen Mary University of London. He completed a Master's in electrical engineering at MIT in 2005 and a PhD in music composition at the University of Pennsylvania in 2009. His research focuses on new musical performance interfaces, particularly augmented instruments which extend the capabilities of familiar acoustic instruments.

Andy Farnell

Email: *padawan12@obiwannabe.co.uk*

Andy Farnell is a computer scientist from the UK specialising in signal processing and synthesis. Pioneer of procedural audio and the author of MIT textbook "Designing Sound", Andy is visiting professor at several European institutions and consultant to new media and games companies. He is an enthusiastic advocate and hacker of Free software, who believes in educational opportunities and access to enabling tools and knowledge for all.