Mapping and Interaction Strategies for Performing Environmental Sound

Christian Heinrichs*
Centre for Digital Music
Queen Mary University of London

Andrew McPherson†
Centre for Digital Music
Queen Mary University of London

Abstract

While the design of computational audio models for real-time generation of sound has been gaining increasing attention in the field of virtual reality and games over the last few years, questions related to expressivity and human performability have remained largely unexplored. Unlike in the design of interactive sonic artefacts a performable model requires a different approach to parametrisation and interaction. A model of a squeaking door is presented along with three contrasting mapping strategies between a generic touch-based interface and parameters controlling phenomenologically meaningful sound qualities. Each of these mapping strategies is evaluated in a controlled study based around a set of four metrics proposed by the authors. Correlations between quantitative and qualitative data verify the evaluation procedure for each of these metrics.

1 Introduction

Human performance is still a relatively unexplored topic in the context of computational audio models for immersive environments and games. While fields such as interactive sonification, mapping strategies for physical models and the rapidly expanding field of sonic interaction design have benefited from a vibrant and growing research community over the last two decades, interactive synthesised sound has only recently started to become a topic of interest in the field of game audio and virtual reality.

In the work outlined in this paper we seek to answer the following question: given a computational sound model, what is the best mapping strategy which allows a user to perform believable environmental sounds? Creating a performable sound model requires a different set of design targets from those involved in the design of an interactive sonic artefact. The interaction requires precise control over linearly independent parameters within a space of phenomenological sound qualities.

We have designed a model of a squeaking door on the premise that it would require control inputs existing outside the realm of physical parameters. Three mapping layers are presented, which can be used to interface a simple touch-based controller to the model. While two of these are based on simple 'one-to-one' and 'many-to-many' principles the other falls within a less common class of mapping strategies that we refer to here as physically-inspired control layers (PhICLs). Three potentially quantifiable metrics borrowed from the field of digital musical instrument design are presented for the evaluation of a performance interface for computational audio models, along with the fourth metric of 'believability', which is crucial for the evaluation of environmental sound within an immersive or narrative context. Finally, results from a user evaluation study are presented, in which each of the four metrics are tested for the three control layers.

1.1 Designed behaviour in computational models

Within the context of multimodal immersive environments, the majority of computational approaches share a dependency on a separate *physics engine*, i.e. a computational layer that calculates the behaviour of all audio-visual objects in the scene using complex algorithms. This is useful for some scenarios, for example where quantifiable information needs to be extracted from the sound image, or simply when audio-visual synchronicity is required. However, in many other audio-visual scenarios - particularly those following some form of narrative - the sound may be required to contradict the image and indeed physical laws themselves. Chion refers to 'audio-visual phrasing' [5] for the way in which sound and image can support or contradict each other over time. The function of this is to guide the listener's attention to details that would otherwise not be present in the image, amounting to what he refers to as 'added value' in audio-visual media. Such counterpoint is difficult - or inefficient at best - to achieve with a 'hard-coded' behavioural layer (including the algorithms that make up a typical physical model but also most approaches to parametrisation in physically informed [8][13] and so-called 'hybrid' models [2]). A different means of encoding behaviour into the computational model would be desirable in these instances.

1.2 Ergo-audition vs Performed Sound

The crucial difference between a model designed for human performance and one to be driven by simulated physical processes is in the way the model's behaviour is collapsed into variant (or 'behavioural') parameters. Distinguishing between 'ergo-audition' and 'performed' sound is a helpful way of marking this difference.

Chion's term 'ergo-audition' refers to the sounds of one's own actions in the world [6]. In an immersive environment these would include virtual footsteps accompanying the avatar's movement, sonification of limb and head movements, engine sounds when driving a vehicle, and so on. While the user or player (ideally) has real-time control over the variation of these sounds, the interaction is fundamentally different from performing sounds as part of the design process: the parameter space has already been collapsed into a few dimensions that describe a specifically designed behaviour. More specifically, the player (or the avatar through the player) needs to perform a causal gesture (e.g. walking for footsteps, accelerating or braking for engine sounds) in order to produce a sound. Caramiaux et al [4] and Godøy [15] distinguish these types of gestures that mimic a sound's cause from those that trace its inherent morphology. While both types of gestures could be said to be 'performative' the latter type is likely to be more suitable in creating complex sound trajectories as part of a design task, as it provides an easier means for controlling the sound along abstract timbral dimensions.

Hug acknowledges the possibility of employing performative strategies in the design of interactive sonic objects [16]. He refers to the mediated nature of immersive environments as a 'second order' experience, noting an element of re-enactment when interacting with sounds in a virtual environment (i.e. ergo-audition through the actions of an embodied avatar). This affords a different set of design strategies than methods typically used in the field of Sonic Interaction Design (SID), which focuses on guiding or designing interactions with everyday objects (or 'commodities') through the use of interactive sound.

The suggestion is that to *perform the behaviour* of a sound model (as opposed to 'activating' it through embodied interaction) we require a parameter space that enables more abstracted control over the model's sound generation, such that the behaviour of a potential sounding object can be extracted performatively from a larger timbre space.

*e-mail: c.heinrichs@qmul.ac.uk
†e-mail: a.mcpherson@qmul.ac.uk

2 Related Work

2.1 Re-purposed physical objects

The majority of the work dealing with the performance of environmental sound falls into the category of re-purposed physical objects. That is, where the interaction affordances of one sound-producing object are used to control another sound generating mechanism that is related in one way or another. Essl and O'Modhrain have created a variety of objects with clearly associated affordance structures, such as a box of pebbles or a whiteboard eraser, to control parameters of physical or physically informed models [11] [10]. Cook uses an accordion to control the complex multi-dimensional parameter input of a voice synthesis model [7]. A variety of other controllers for his physically informed sonic models (PhISMs) [8] are also described in [7]. In [3] control input from the saxophone is used to control a complex physical model of a bowed string. This process, referred to as instrumental controller substitution [25] allows new and meaningful ways of timbre exploration for physical models without the need to re-parametrise the model.

2.2 Independent mapping layers

Other work focuses more specifically on an independent mapping layer that exists between the physical controller and the sound model. Bevilacqua et al [1] have developed the Gesture Follower, which uses dynamic time warping and hidden markov models to recognize trained gestures and precisely follow their execution and variation over time. This work is extended further by Francoise et al [14] to include real-time segmentation of gestures, enabling even greater control over sound parameters using abstract gestures that are meaningful to the given user/designer. Momeni and Henry [19] propose the use of a fully independent control layer to control audio and/or video synthesis. This is achieved by actuating simulated physical systems such as mass-spring models using gestural input and connecting the output of the control layer to synthesis parameters in some meaningful way. Although no particular type of synthesis is specified here, this approach can be applied to the control of physical or physically inspired sound models. Such an implementation was carried out recently by Thoret et al [26] using a model of a bowed string as a proxy generating parameters for a friction-based sound model. Coincidentally, we have implemented a similar control layer for our own squeaky door model as one of three mapping strategies evaluated below.

3 Four Metrics for the Performance of Environmental Sound

Currently there exist no established means to evaluate how well an interface lends itself to the human performance of a given environmental sound. Borrowing from the field of digital musical instrument design we propose four aspects of a performable model that can be used as a foundation for suitable metrics within such an evaluation task.

3.1 Micro-Diversity and Mid-diversity

In [18] Jordà describes three levels of 'diversity', which are presented as a new take on recurring concepts within the field of digital musical instrument design previously referred to as the 'versatility' or 'flexibility' of an instrument [20]. The first category, macro-diversity, refers to the ability of an instrument to perform in different contexts or styles. For example, a guitar has a higher level of macro-diversity than an oboe because it can applied within a wide range of musical styles and public contexts. The next category is mid-diversity, which is most succinctly described as how distinguishable two different pieces played on the same instrument can be. An instrument with very low mid-diversity will sound as though it is always playing the same piece of music. Finally, micro-diversity is the instrument's ability to control subtle nuances of a musical sound - in other words, how different two performances of the same piece can be.

While these three metrics were specifically conceived to describe musical instruments, they are equally relevant to the expressive performance of environmental sounds. Thus, micro-diversity refers to the ability to control subtle nuances of a sound - for example, does the sound of two swords scraping along each other always sound the same, or can various parameters be controlled by the interface? Mid-diversity refers to the behavioural breadth that the performable model can produce: can the sound produce every sound associated with a sword, or only the sound of two swords scraping each other? Macro-diversity is less easily translatable to environmental sound. The likeliest interpretation would be the amount of different objects that can be represented by the performable model (much in the same way a foley artist might use a 'prop' to perform the sound of an unrelated object).

3.2 Repeatability

Controllability has long been recognized as an important metric in the evaluation of musical interfaces [27] but is equally applicable to the performance of non-musical sounds. It refers to the ability to maintain control over features and timing of a musical trajectory and is equatable to 'repeatability', which is the preferred term that will be used here. This metric can easily be evaluated by testing the ease of repeating a sound with the same interface.

3.3 Believability

While expressivity is an undeniable (and also unquantifiable) aspect of any human performance, environmental sound brings with it a further metric, which is crucial to its function in immersive and narrative contexts. This is namely that a sound can only be associated with an object or environment if it is believable as such. Note that believability is *not* the same as realism, which is why this aspect is metric is not testable by purely physical or mathematical means (e.g. similarity to a signal). A human listener or interactor is required to believe a heard sound to be the intended part of a given environment. Therefore believability should ideally also be tested as part of the listener's experience *within* its intended context or environment.

4 Development of a Performable Squeaky Door Model

4.1 Reasoning behind Choice of Model and Controller

With the aim of studying the incorporation of human performance into the design of computational audio models we sought to develop a sound model that had high potential for expressive variation temporally, timbrally and contextually. We chose to develop a model of a squeaking door as a case study for the following three reasons:

1. The sound is not restricted to any particular morphology (e.g. as opposed to impact or explosion sounds which always consist of a rapid attack and a gradual decay toward silence).

2. The sound can have a lot of timbral variation: squeaky doors can groan, screech, jitter, etc.

3. The causal relationship between timbral variation and physical mechanism (i.e. motion around the hinges) is opaque to the casual listener. This means that when presented with the task of producing the sound for the same physical door two performers are likely to generate different squeaks depending on the context of the sound effect and personal experience.

For similar reasons we decided to use a generic controller - a trackpad capable of sensing two axes and touch area - as a means of physical input. This enabled us to perform a controlled study of mapping layers between the physical controller and the sound model, while sticking to an interface that is widely-used and likely to be familiar to the participant.

4.2 Development of a Hybrid Model through Phenomenological Analysis

The well-known sound of a squeaking door is caused by stick-slip friction, a dynamic process caused by unstable interactions between frictional and applied forces when an object is moved along a surface. Historically, stick-slip friction has proven notoriously difficult to simulate algorithmically. A comprehensive overview of physical models is provided by Serafin [24] who also developed a realistic synthesis model, available as PureData and Max/MSP externals and also as part of the Sound Design Toolkit [9].

While physical approaches are capable of generating highly convincing squeaking and creaking sounds we found that they were not suitable for this study. The control parameters are, naturally, biased towards interfaces or objects that are associated with friction sounds such as a bowed string or an actual door on a set of hinges. The remaining invariant parameters affect the dynamic behaviour of the model and, being based on a particular complex model of stick-slip friction, are highly unpredictable to the casual listener. Furthermore we were interested in the process of developing a model from the ground up with human performance (that is not restricted to causal gestures) in mind. Instead of a physical model, we developed a physically inspired model using a mixture of physical and phenomenological analysis.

Starting with an impulse train passed through a bank of delay and filter-based resonators (similar to the model outlined by Farnell in [12]) we used an iterative process of analysis and implementation to extend the model. We first acquired recordings of real doors, extracted from sample libraries as well as own recordings. Through close listening we identified independent morphological features. If this feature was not yet capable of being produced by the model it would be implemented by inserting the necessary signal processing blocks along with an associated parameter. For example, while the pitch of the squeak was controllable by setting the frequency of the impulse train, we also found that the pitch was often subject to noisy micro-variations, and thus a separate 'roughness' parameter was implemented to recreate this effect independently of the pitch. This process was repeated until the model was capable of reproducing our acquired recordings to a satisfying degree of accuracy. Note that while knowledge of the physical cause behind the sound is valuable in this process of extending the parameter space, the analysis involved is better described as phenomenological - listening for variations in the way something *sounds* rather than in the way something *works*. The resulting model can be classified as a hybrid between a physically informed and an abstract synthesis model [2] as it is capable of mimicking a wide variety of physically generated sounds, yet does this through a static signal chain controlled by a linearly independent set of non-physical parameters (not unlike a subtractive synthesiser).

The final model was extended from a single variant parameter to a total of 17 variant parameters and more than 50 invariant parameters. The large parameter space was simply due to the large library of targeted sounds (e.g. some recordings featured doors with two squeaking hinges rather than one, so variant parameters had to be doubled at one point of the analysis stage). It was nonetheless possible to generate a wide breadth of sounds by changing only two or three of the variant parameters simultaneously over time.

4.3 Mapping strategies

Three control layers were developed for analysis in this study. Each of these was based on the same set of input parameters from the physical controller (vertical position, velocity, touch size) and output parameters to the sound model (pitch, roughness, impulse shape).

4.3.1 One-to-one mapping

In this approach, as the name suggests, each input parameter has independent control over a synthesis parameter. Vertical position controls pitch, where higher pitches are produced when placing a finger at the top of the touch surface. Touch size controls roughness, where a larger value results in a higher amount of roughness. Finally, velocity controls the brightness of the sound (i.e. the attack component of the impulse train), where a faster velocity results in a duller sound (i.e. smoother attack).

4.3.2 Physically inspired control layer

Drawing from implementations outlined in section 2.1 we designed a mapping based on a physical control metaphor, which we refer to here as a Physically Inspired Control Layer (PhICL). The control

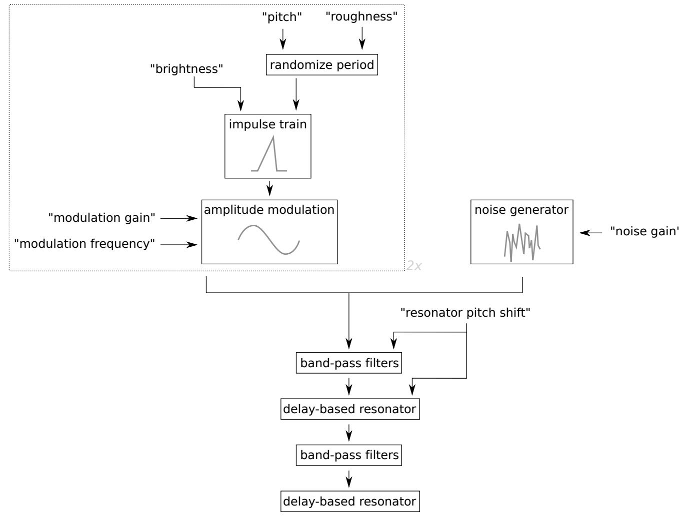

Figure 1: Block diagram of squeaky door model.

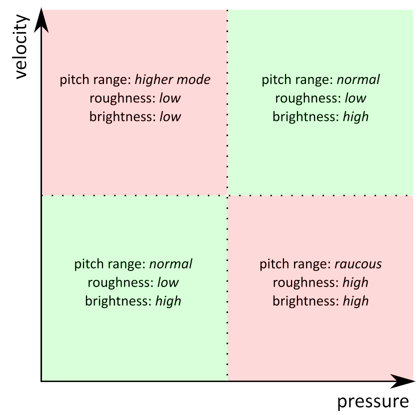

Figure 2: Relationship between virtual bow pressure and velocity and synthesis parameters controlling door squeak model

layer is loosely based on the interaction between a bow and a string and can be seen as an independent component placed between the physical control input and the sound model. As mentioned above, the physical mechanism behind timbral variations of door squeaks is relatively opaque making it difficult to integrate known causal relationships into the control layer. On the other hand, a bowed string produces a similar range of sounds through comparatively transparent relationships between input and output parameters. Depending on the combination of bow velocity and pressure it is possible to produce 'raucous' (i.e. aperiodic and harsh) squeaks, 'normal' periodic sounds, or higher modes of the given string (a detailed overview is provided by Schelleng [23]). The 'fundamental' pitch is controlled by varying the length of the resonating segment of the string (e.g. moving the left hand along the neck of a violin). The resulting sounds can be expressed using the three chosen synthesis parameters. As illustrated in Figure 2 the sounding 'sweet spot' can be achieved by applying the right combination of pressure and velocity. Additional non-linearities were added to the model such as hysteresis between pitch ranges, loosely emulating the effect of stiction.

Finally, physical control parameters are mapped to the PhICL's input parameters. Velocity and touch size are directly mapped to the velocity and pressure parameters of the control layer. Absolute vertical position is mapped to the fundamental pitch of the model,

	Vertical pos.	Velocity	Touch size
Pitch	0.6	0.2	0.2
Roughness	0	0.2	0.8
Brightness	0	0.5	0.5

Table 1: Weightings of input parameters in convergent mapping strategy

maintaining the orientation from the previous one-to-one mapping (i.e. up for a higher pitch).

4.3.3 Arbitrary convergent mapping

If a one-to-one mapping strategy is the most simple way of mapping control to synthesis parameters, a convergent mapping approach lies at the opposite pole of mapping complexities suggested by Rovan et al in [22], now more commonly referred to as many-to-one or many-to-many mapping [17]. Here each synthesis parameter is controlled by more than one single control dimension. While the PhICL can also be understood as a convergent mapping, this layer is a more straight-forward implementation in that control parameters are simply added together (after being normalized and individually weighted) to produce an output parameter. To ensure a minimum level of playability each input parameter was assigned a different weighting based roughly on the relationship between parameters in the PhICL (see Table 1).

5 Experimental Procedure

The study consists of two parts. The first part is a performance study in which the participant is asked to perform a variety of squeaky door sounds according to a pre-defined set of tasks. The second part is a listening study in which doors performed using one of the interfaces must be distinguished from door squeaks that were modelled (using the same sound model) on recordings of real doors.

5.1 Environment and Set-up

The experiment is carried out in a controlled, quiet environment and the participant uses headphones throughout. Tasks are given to the user through an automated graphical environment that is navigated using a specially labelled keyboard, allowing each participant to carry out the experiment without the need for supervision. Nonetheless, the participant is given the opportunity to ask questions at any point throughout the study should any of the instructions be unclear. In the first part of the study sounds are performed on the trackpad which is attached to the table in front of the keyboard. The participant is told that they can only use one finger to generate sounds (to avoid the use of common 'swipe' or 'pinch' gestures).

5.2 Part 1

In the first part of the experiment the participant is sequentially presented with nine imaginary scenarios for which he/she is asked to perform the sound of a squeaky or creaking door. Each scenario consists of 2-4 sentences written in evocative language to encourage maximum variation for each performed sound effect, for example:

"Two shifty characters are closing a deal in a secret room in the basement of a bar. As one of them hands a briefcase over the door unexpectedly opens. Both characters instantly pull guns out of their pockets and aim at the door, looking at each other suspiciously. Perform the sound of the door."

Invariant parameters of the sound model as well as convolution reverb settings are loaded as pre-defined presets for each scenario corresponding to the type of door and space specified by the scenario. Before being presented with the first scenario the participant is given a tutorial on the current mapping strategy (referred to in the study as an 'interface'), which is chosen randomly. The tutorial consists of no more than four simple images accompanied by text demonstrating how finger movements affect the sound quality (e.g. "Large touch area and low movement speed results in a low-pitched unstable squeak"). The participant is allowed to practise while navigating through the tutorial. The participant is then given the first task scenario and can spend as much time as he/she wants practising before being asked to record the targeted sound effect three times. His/her favourite performance is then chosen (given the opportunity to listen back to each one). The participant is then asked to repeat the same sound five times, attempting to match the original one as accurately as possible. Following this he/she is asked to repeat the sound again at a higher or lower pitch, or a faster or slower speed (chosen randomly for each scenario).

The procedure is repeated for each scenario. After three performance tasks have been completed a new control layer is enabled and the participant is given the corresponding tutorial before proceeding to the following tasks.

5.3 Part 2

In the second part of the experiment the participant listens to a radio play that was specially written and produced for this study. [1] The radio play is 14 minutes long and centres around two characters viewing an old house that used to belong to one of the character's relatives. All dialogue and additional foley tracks were treated with convolution reverb corresponding to each room that is visited by the characters. The play contains nine occurrences of squeaky doors, windows and gates, which physically correspond to the ones featured in the first part of the study (e.g. 'iron gate', 'stiff window', 'light wooden door', etc.). While invariant parameter settings are the same as in their corresponding performance task scenarios, gain and reverb settings are adjusted to blend in naturally with the sound image of the radio play.

For each participant 6 out of 9 randomly selected door sounds are taken from favourite performances of the last participant (for the corresponding door type). The remaining 3 have been modelled on recordings of real doors (extracted from sample libraries and own recordings). The recordings were matched by trial-and-error using an offline sequencer that stored and sent OpenSoundControl messages to the model. Note that the doors were matched using no other variant parameters than the three that were controllable using the trackpad interface.

The participant is explained the difference between a 'performed door' and a door that was modelled on a pre-existing recording before listening to the radio play. After each time that a door sound occurs in the radio play the participant is asked to choose whether the sound effect was performed or based on a real recording by pressing a corresponding key. This decision is encouraged to be made as the play continues (pausing after five seconds of inactivity) - a flashing screen prompts a quick response by the participant.

5.4 Data Collection

Control and synthesis parameters are logged throughout the first part of the study. Logs are trimmed and saved to individual files for each trial and repeat to ease later analysis.

Additional qualitative data is collected by means of a survey consisting of a mixture of Likert-style ratings, yes/no questions and opportunities to leave further comments.

5.4.1 Participants

15 people (7 female, 8 male) aged between 27 and 40 participated in the study. 9 people had extensive audio experience or played a musical instrument for longer than five years. The remaining people had some experience editing or recording sound.

6 Results

6.1 Repeatability

The relative success of repetition across interfaces was measured by applying Dynamic Time Warping (DTW) between repeated sequences and their corresponding original sequences. The results are shown in Table 2.

Results demonstrate a high correlation with qualitative data obtained from the survey. Responses by participants to the statement

[1] Written and performed by David G Lees and Stefanie Ritch. Recorded and produced by Christian Heinrichs at studios provided by the Centre for Digital Music, Queen Mary University of London, UK.

	Rel. rating	Mean Alignm. Cost	S.D.
One-to-one	2.294	3.858	9.720
Many-to-many	1.000	1.682	4.726
PhICL	3.340	5.719	8.122

Table 2: Results from repeatability test

	Quantitative	Qualitative
One-to-one	2.294	2.625
Many-to-many	1.000	1.000
PhICL	3.340	7.000

Table 3: Correlating relative repeatability ratings for quantitative and qualitative measurements. Lower number corresponds to a higher success rate.

'I found it easy to repeat the sounds during the repetition task' on a Likert scale made it possible to generate a corresponding qualitative relative repeatability rating (see Table 3).

6.2 Micro-Diversity

Micro-diversity (the ability to control subtle nuances of a sound) was measured using the extended repetition task in which participants were asked to repeat a sound that they previously performed at a higher or lower pitch or at a faster or slower speed. By discounting pitch and time differences and testing divergence between original and repeated tasks we are measuring how easily a person can vary one aspect of sound without affecting the rest of the parameter space. A quantitative rating was obtained as well through Likert-ratings to the statement 'The interface gave me high degree of control over subtle nuances of the sound (e.g. while refining the sound for a given scenario)'.

DTW was applied again to measure similarity between original and repeated synthesis parameters. Beforehand, pitch parameters in repeated sequences were vertically aligned to the corresponding originals in order to discount intentional differences resulting from the pitch-variation tasks. Results show lower alignment costs for the many-to-many mapping and much greater divergence for the PhICL (see Table 4).

Quantitative measurements and subjective responses correlate to the extent that one-to-one and many-to-many mappings were rated higher than the PhICL; however, the one-to-one mapping was perceived to allow greater control over nuances than the many-to-many interface (see Table 5).

6.3 Mid-Diversity

Mid-diversity is less straightforward to measure by quantifiable means. The first observation worth noting is the distribution of performable parameter space for each of the mappings. Figure 3 contains three scatter plots of all the performed sequences (discounting repeats) for each mapping strategy (colours correspond to performance tasks corresponding to sequence data). While the one-to-one mapping has the least biased parameter space, the PhICL seems much more restricted by comparison. This is to be expected. The physically inspired control layer is purposefully designed in such a way that any given door sound can have high roughness and brightness values (when within the 'racous' or 'normal' pitch ranges) or low roughness and brightness values (when within the 'higher mode' range) but no other combination of the two parameters.

Mid-diversity is not a measure of the parameter space itself but of how much performed sounds tend to vary in different contexts. In the case of this study this means measuring how much the performed trajectories vary across different tasks for each interface. In an attempt to test this we fitted Gaussian Mixed Models (GMM) with ten weighted components to the parameter data of each task. For each task the log probability of each data point was then calculated under the GMM model of every other task. Finally, normalized mean log probabilities were calculated for each interface and used to generate distance metrics (see Table 6).

Subjective responses were additionally gathered by prompting a Likert-rating for the statement 'Overall the interface was capable of producing a wide breadth of door sounds'. Ratings show high correlation to the quantitative results, favouring the one-to-one mapping. Normalized responses are shown in the first column of Table 6.

	Rel. rating	Mean Alignm. Cost	S.D.
One-to-one	1.583	4.6025	8.1401
Many-to-many	1.000	2.9082	9.5846
PhICL	2.799	8.1401	7.5321

Table 4: Results from micro-diversity test. Lower number corresponds to a higher success rate.

	Quantitative	Qualitative
One-to-one	1.583	1.000
Many-to-many	1.000	1.130
PhICL	2.799	1.857

Table 5: Correlating relative micro-diversity ratings for quantitative and qualitative measurements. Lower number corresponds to a higher success rate.

	Subj. Responses	Mean Distance	S.D.
One-to-one	1.000	1.222	0.687
Many-to-many	0.871	0.965	0.641
PhICL	0.677	0.282	0.255

Table 6: Qualitative and quantitative results for Mid-Diversity. Higher number corresponds to a greater measure of diversity.

6.4 Believability

The results from the listening study are shown in Table 7. In addition to the data gathered during the listening study we also collected further subjective responses to likert-style questions in a separate questionnaire.

According to the questionnaire results door sounds were generally perceived as believable regardless of whether they were thought to be performed or not. This is reflected in the low percentage of correct responses for the doors based on real recordings. On the other hand, door sounds performed using the many-to-many mapping were easier to distinguish as being performed compared to all other sounds.

6.5 Summary

Participants consistently preferred the many-to-many mapping in the questionnaires and this also had the highest ratings for repeatability and micro-diversity, however these sounds were the least realistic in that listeners could more readily distinguish them from sounds based on recordings. This suggests that, potentially unlike musical instruments, the reaction of the performer might not be the best (or certainly not the only) metric in evaluating the success of the interface.

The consistent low ratings of the PhICL and comparative success of the many-to-many mapping may seem somewhat surprising considering the intuitiveness and controllability that is often associated with physical control mechanisms. On the other hand, ratings are likely to change after longer exposure to the interface. For in-

	% Responses correct	Total responses
Based on recordings	44.4%	45
One-to-one	56.3%	32
Many-to-many	69.0%	29
PhICL	55.2%	29

Table 7: Percentage of correct listening responses for each category.

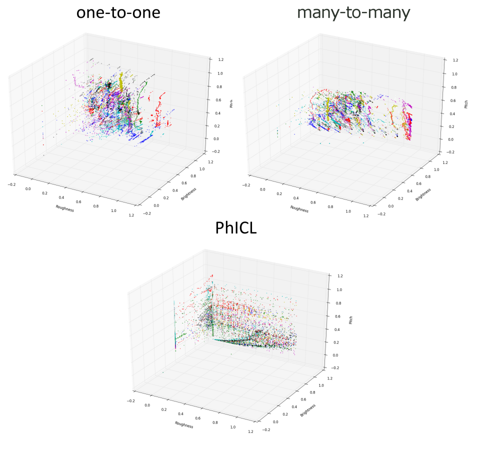

Figure 3: Scatter plots of all performance data for each interface. Colours correspond to different tasks.

stance, the PhICL might yield much higher repeatability and micro-diversity ratings after participants have spent more time practising performing sounds on the instruments. The only way to map the learning curve of a performance interface is through a longitudinal study [18] [21]. On the other hand it is important to remember that, unlike musical instruments, performable interfaces for computational audio models require a very steep and low learning curve if they are to be used as part of the design process.

7 Conclusion

We have presented a performable model of a squeaking door . Performable models require a different approach to parametrisation than most interactive sonic artefacts. A means of controlling the sound along dimensions of sound quality is required and therefore linearly independent parameters affecting phenomenologically meaningful aspects of the sound may be more suitable than physical parameters affecting dynamic behaviour associated with the represented object.

Using our proposed set of four metrics - micro-diversity, mid-diversity, repeatability and believability - we evaluated three mapping strategies developed to control our physically-informed model of a squeaking door. While, surprisingly, an arbitrary convergent mapping strategy had the highest ratings in both quantitative analysis and subjective responses, believability ratings gathered from a listening study suggest that player-centric evaluation may not be enough to evaluate the suitability of a performance interface. Correlations between subject questionnaires and quantitative data show that our quantitative approach to the evaluation metrics is indeed suitable. While a longitudinal approach would eliminate potential biases associated with the learning curve of the interface (e.g. initial unfamiliarity), the immediacy of a performable sound model is important if it is to be feasibly implemented as part of a larger design process.

Acknowledgments

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

References

[1] F. Bevilacqua, N. Schnell, N. Rasamimanana, B. Zamborlin, and F. Gudy. Online gesture analysis and control of audio processing. In *Musical Robots and Interactive Multimodal Systems*, page 127142. Springer, 2011.

[2] S. D. Bilbao and Wiley InterScience (Online service). *Numerical sound synthesis finite difference schemes and simulation in musical acoustics*. John Wiley & Sons, [Hoboken], 2009.

[3] M. Burtner. The metasaxophone: concept, implementation, and mapping strategies for a new computer music instrument. *Organised Sound*, 7(02), Jan. 2003.

[4] B. Caramiaux, P. Susini, T. Bianco, F. Bevilacqua, O. Houix, N. Schnell, and N. Misdariis. Gestural embodiment of environmental sounds: an experimental study. In *Proceedings of the International Conference on New Interfaces for Musical Expression. AR Jensenius et al., Eds*, 2011.

[5] M. Chion. *Audio-Vision: Sound on Screen*. Columbia University Press, 1994.

[6] M. Chion. *Le Son*. Fac. Srie Cinma et image. Nathan universit, 1998.

[7] P. Cook. Principles for designing computer music controllers. In *Proceedings of the 2001 conference on New interfaces for musical expression*, page 14, 2001.

[8] P. R. Cook. Physically informed sonic modeling (phism): Synthesis of percussive sounds. *Computer Music Journal*, 21(3):3849, 1997.

[9] S. Delle Monache, D. Devallez, C. Drioli, F. Fontana, S. Papetti, P. Polotti, and D. Rocchesso. *Sound synthesis tools for sound design*. Deliverable, 2008.

[10] G. Essl and S. O'Modhrain. Scrubber: an interface for friction-induced sounds. In *Proceedings of the 2005 conference on New interfaces for musical expression*, page 7075, 2005.

[11] G. Essl and S. O'Modhrain. An enactive approach to the design of new tangible musical instruments. *Organised Sound*, 11(03):285, Nov. 2006.

[12] A. Farnell. *Designing Sound*. Mit Press, 2008.

[13] A. Farnell. Behaviour, structure and causality in procedural audio. In *Game Sound Technology and Player Interaction: Concepts and Development*. Information Science Reference, 2011.

[14] J. Francoise, B. Caramiaux, and F. Bevilacqua. A hierarchical approach for the design of gesture-to-sound mappings. In *Proceedings of the 9th Sound and Music Computing Conference (SMC)*, 2012.

[15] R. I. Gody. Images of sonic objects. *Organised Sound*, 15(01):54, Mar. 2010.

[16] D. Hug. Investigating narrative and performative sound design strategies for interactive commodities. In *Proceedings of the 6th International Conference on Auditory Display*, CMMR/ICAD'09, page 1240, Berlin, Heidelberg, 2010. Springer-Verlag.

[17] A. Hunt and M. M. Wanderley. Mapping performer parameters to synthesis engines. *Organised Sound*, 7(02), Jan. 2003.

[18] S. Jord. Instruments and players some thoughts on digital lutherie. *Journal of New Music Research*, 33, 2005.

[19] A. Momeni and C. Henry. Dynamic independent mapping layers for concurrent control of audio and video synthesis. *Computer Music Journal*, 30(1):4966, 2006.

[20] K. Nishimoto, C. Oshima, and Y. Miyagawa. Why always versatile?: Dynamically customizable musical instruments facilitate expressive performances. In F. Thibault, editor, *NIME*, pages 164–169. Faculty of Music, McGill University, 2003.

[21] S. O'Modhrain. A framework for the evaluation of digital musical instruments. *Computer Music Journal*, 35(1):2842, 2011.

[22] J. B. Rovan, M. M. Wanderley, S. Dubnov, and P. Depalle. Instrumental gestural mapping strategies as expressivity determinants in computer music performance. In *Proceedings of Kansei-The Technology of Emotion Workshop*, page 34. Citeseer, 1997.

[23] J. C. Schelleng. The bowed string and the player. *The Journal of the Acoustical Society of America*, 53:26, 1973.

[24] S. Serafin. *The sound of friction: real-time models, playability and musical applications*. PhD thesis, stanford university, 2004.

[25] S. Serafin, M. Burtner, C. Nichols, and S. OModhrain. Expressive controllers for bowed string physical models. In *Proc. DAFX 2001, Limerick, Ireland*, 2001.

[26] E. Thoret, M. Aramaki, C. Bourdin, L. Bringoux, R. Kronland-Martinet, and S. Ystad. Synchronizing gestures with friction sounds: work in progress. In *Sound, Music & Motion-Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)-15-18 oct. 2013-Marseille, France*, page 190199, 2013.

[27] M. M. Wanderley and N. Orio. Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal*, 26(3):6276, 2002.