Evaluating and Preparing Object-Contact Models For User Interaction

Max Ardito
MUMT 618

December 11, 2022

1 Introduction

The physical modeling of contact between rigid objects is quite unique in comparison to more traditional physical audio systems that model things like acoustic instruments and reverberant rooms. A system that models the sonic interaction between two objects must simultaneously characterize the physical system in question while also accommodating the wide array of hypothetical rigid objects, types of contact (scraping, rolling, bouncing), and object geometries and materials. For instance, the spectral properties of a bouncing basketball differ greatly to those of a glass bottle rolling, or a pencil scrapping on a sheet of paper.

In the corpus of relevant literature, various methodologies have been prescribed to the problem. One family of approaches attempts to model these systems by boiling down the sum total of interactions, geometries, and materials to a few fundamental parameters. This approach relies heavily on source-filter models and analysis/resynthesis techniques, often starting with a recording of the interaction of the objects in question and then extracting data pertaining to the resonance and excitation of the interaction. This data can later be reconstructed and encoded back into the original sound. A parametric approach such as this is best demonstrated in Lagrange et al. [1].

An additional fine-tuned approach of the source-filter model is presented in Lee et al. [2] which focuses specifically on modelling ball rolling sounds. While constrained to a more specific family of object interactions, this model uses a more accurate approach by encoding the gradual change in excited and supressed modes over time.

In this paper, a survey of the techniques used in Lagrange et al. and Lee et al. will be presented, followed by an evaluation of the modal analysis/resynthesis techniques. This survey will lead to an attempt at constructing a model based on the paramterized generation of granular microcollisions using the aforementioned modal analysis/resynthesis techniques. Beyond these findings, a schema for mapping the updated model to accomidate real-time human interaction using a simple tablet or touchscreen device is proposed. Other physics-based models are mentioned in the conclusion.

2 The Object-Agnostic Approach

The approach described in Lagrange et al. provides a method in which the system in question is not constricted to a certain family of objects or interactions. Instead, a combination of HR analysis methods [3] and envelope modeling techniques are used to parameterize an inclusive model of sustain object contact that is agnostic to object geometries, materials, and interaction types (scraping, rolling,

bouncing). The highest level overview of the proposed model uses a recorded signal x of the interaction in question and generates a reconstructed signal $\hat{x}$ made up of the following components...

$$\hat{x} = d * i * s \tag{1}$$

where d is a series of amplitude modulated dirac functions, i is an ideal impact or "microcollision" extracted from x used to parameterize an ideal excitation envelope, and s is an series of N second-order filter sections whose coefficients contain information about the resonant characteristics of the objects in question. This model thus treats all of the resonances obtained from the recording as a single filter s and all of the excitation patterns that excite this filter as a source $d * i$.

In terms of a continuous time physics-based analysis, the approximation that is to be obtained when extracting s will perform a sort of modal analysis of the objects in question. When an object is vibrating freely, it will resonate at various modal points governed by the object's geometric and material properties [4].

$$\mathcal{M} = \{f, d, \mathcal{A}\} \tag{2}$$
$$d * i \rightarrow \{f_{1\times N}, d_{1\times N}\} \tag{3}$$
$$s \rightarrow A_{N\times K} \tag{4}$$

In (2) [4], a modal model M is approximated using a superposition of second order filters [5]. In the context of the Lagrange et al. model, the source signal $d * i$ encodes information about the amplitude and phase of the modes and the filter s encodes their frequency and damping factors f and d [1]. This modal model can then be approximated using the impulse response $y(t)$ given in (5) [4].

$$y(t) = \sum_{n=1}^{N} a_{nk} e^{-d_n t} \sin(2\pi f_n t) \tag{5}$$

The question then becomes what the best way might be to extract these N modes from the source signal. The authors propose the use of high-resolution (HR) methods in order to extract the precise sinusoidal components within an ideal window of the recorded interaction. HR analysis techniques attempt to overcome the resolution issues found in Fourier analysis by utilizing properties of the signal subspace [3]. The method found by Lagrange et al. to be most apt in analyzing the sinusoidal components of a rolling object was the ESPRIT algorithm, however the choice of an ideal HR method was seen to be somewhat hollistic, as a number of extra preprocessing steps were required to obtain accurate filter parameters.

ESPRIT performs an eigenanalysis of the autocorrelation matrix, extracting the K-highest eigenvalues, corresponding to the first K modes of the model [1]. These modes form the basis for the "signal subspace" and can be modeled by a series of second order filters, matching the K modes to the first z_k filter parameters. Performing ESPRIT is not unlike other subspace techniques involving matrix dimensionality reduction, such as principal component analysis (PCA) or singular value decomposition (SVD). This is demonstrated in (6) [6], which presents the end product Φ of the ESPRIT approximation. The eigenvalues of the Vandermonde matrix Λ can be extracted as complex conjugate sections.

$$\Phi = Q\Lambda Q^{-1} \quad | \quad \Lambda = \begin{bmatrix} z_1 & 0 & \cdots & 0 & \overline{z_1} & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 & 0 & \overline{z_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_K & 0 & 0 & \cdots & \overline{z_K} \end{bmatrix} \tag{6}$$

A modal representation of the interaction is thus obtained from plugging these coefficients into the series of filter sections. By doing so, we thus approximate the most prominant sinusoidal components of the recorded signal. Furthermore, we can then obtain an adequate approximation of the excitation signal if we deconvolve the resulting filter from the original signal via spectral division in the frequency domain. In other words, if we take the original recorded signal x and perform $STFT\{x(n)\} \rightarrow X(e^{-j\omega})$, we can derrive the excitation signal by performing (8) [1], and then obtaining the time domain excitation signal $ISTFT\{E(e^{-j\omega})\} \rightarrow e(n)$.

$$S(z) = \frac{1}{2}\sum_{k=1}^{K/2} \frac{A_k}{1 - z_k z^{-1} + \frac{A_k^*}{1 - z_k^* z^{-1}}} \tag{7}$$

$$E(e^{-j\omega}) = \frac{S(e^{-j\omega})}{X(e^{-j\omega})} \tag{8}$$

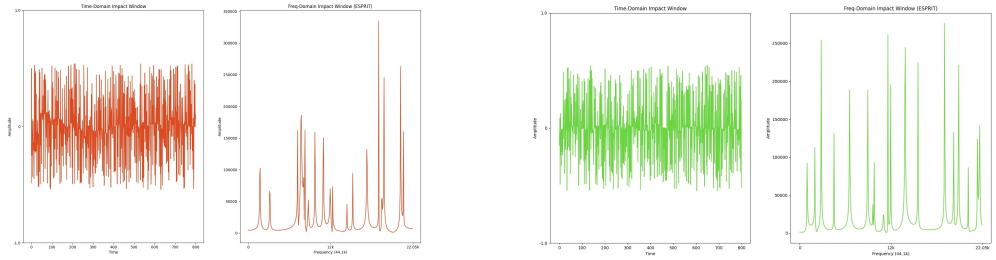

Figure 1: *(Left) Time domain signal and order 20 ESPRIT approximations of a microcollision of a marble rolling on a table. (Right) Same analysis of a microcollision of a coin scraping a table. ESPRIT was performed using the dsatools Python package [7], and second-order filter implementations and spectral deconvolution were done using a MATLAB script written by Gary Scavone [8]. Bandwidth approximation was done in a perceptually hollistic way, as the resonances that characterized these microcollisions are very subtle.*

Now that the filter model have been established, the source or "excitation" must be further parameterized. The goal in this step is to accentuates the slightly uneven contour in a given surface. Since a microcollision can be thought of a sort of onset envelope, Lagrange et al. propose using a Meixner envelope to model the variance captured from collision to collision.

Finally, if the goal is to resynthesize the original recording to evaluate perceputal differences and error metrics, the original signal can be deconvolved once again using the derrived envelopes in order to find a series of m time encoded dirac impulses t_m with amplitude a_m. These encoded impulses can be seen as the set of pairs in (9) [1], each convolved with the Meixner impact envelope.

$$\mathbb{T} = \{t_m, a_m\} \tag{9}$$

While real-time generative modeling is briefly discussed in Lagrange et al., further work can be done on designing a interactive system for the proposed model. Treating this time-encoded set as a statistical distributions that can be sampled from will be discussed further in the subsequent sections.

3 The Object-Specific Approach - Rolling Balls

The object-agnostic approach achieved in the Lagrange et al. paper can be contrasted with a more refined approach that is discussed in Lee et al. [2]. In this case, the set of all possible object collisions is limited to collisions caused by a rolling ball on a surface. Lee et al. presents a similar source-filter model in order to parameterize and reconstruct the original ball-rolling interaction, but with a few important differences.

Perhaps the most important difference is that Linear Predictive Coding (LPC) is used to estimate the the parameters of the filter instead of HR techniques such as ESPRIT. The use of LPC might appear to cause a resolution loss in modal approximation. However, the analysis window on which the LPC is used is not constrained to a single ideal impact such as in Lagrange et al., but rather on a series of multiple windows whose sum total represents all microcollisions found in the recorded signal. This abundance of time-domain analysis windows becomes a crucial in adhearing to the physical properties of the system, as they allow the modes of the vibrating surface to be expressed as a function of time instead of as a static value dissected from a single ideal impact.

The various microcollisions are expressed as K contact sounds $x_k(n)$. They are extracted from the recorded signal by first applying a high pass filter on the signal in order to accentuate the transients of the collisions. These accentuated transients are then used as a tool for onset detection. These onsets are then box-windowed and indexed, each containing unique information about modal excitation over time. The K contact sounds x_k are each analyzed using a 4-band filter bank and downsampled to derive $x_k^{(l)}$ where l is the filter band in question.

$$x_k * \left[\circledast_{l=0}^{4} h^{(l)}\right] \rightarrow X_k \cdot \left[\prod_{l=0}^{4} H_{(l)}\right] \tag{10}$$

We then change basis into the frequency domain and extract each of the 4 subbands' frequency responses, the resynthesized combination of which is shown in (10). Using $X_k^{(l)}(z)$, various information about the modal properties of the system can be derived. Lee et al. note that modes can be both attenuated and suppressed, and that excited modes appear as peaks and suppressed modes appear as time-varying notch patterns [2]. In terms of our source-filter model, these attenuated modes will represent the poles of a hypothetical filter, while the suppressed modes represent the zeros. These time varying spectral peaks (poles) are estimated using LPC while the notches (zeros) are found using a series of spectral operations.

$$N_k^{(l)}(z) = \prod_{m=0}^{M} \frac{(1 - z_{k,m}^{(l)} z^{-1})(1 - \overline{z}_{k,m}^{(l)} z^{-1})}{(1 - \rho z_{k,m}^{(l)} z^{-1})(1 - \rho \overline{z}_{k,m}^{(l)} z^{-1})} \tag{11}$$

The M-order biquad filter in (11) [2] will represent an approximation of the suppressed modes found in $|X_k^{(l)}|$. The parameters of this biquad are obtained by inverting our frequency response to $\frac{1}{|X_k^{(l)}|}$, transforming the M notches into peaks in the magitude representation. Using quadratic polynomial curve fitting, the bandwidth of these M peaks are then estimated in order to represent the zeros as supressed resonances. In terms of bandwidth, these complex conjugate sections $z_{k,m}^{(l)} = e^{\frac{-BW_{k,m}}{2}} e^{-j\omega}$ and the damping factor ρ is given by the authors as 0.95.

$$L_k^{(l)} = \frac{G_k^{(l)}}{1 - \sum_{m=1}^{p_l} a_m^{(l,k)} z^{-m}} \tag{12}$$

Once the zeros are estimated, the poles must be reconstructed in order to obtain the excited modes for each of the contact sounds. This is done by performing an LPC estimate. LPC coefficients $a_m^{(l,k)}$ are estimated and plugged into a transfer function defined in (12). By comparing the coefficients $a_m^{(l,k)}$ to an impulse response $q_k^{(l)}$ obtained from deconvolving (11) from (12), an error metric $e_k^{(l)}(n)$ can be calculated and minimized in order to find the best fit coefficients.

After these approximations have been performed, we can reconstruct each of the l filter sections and furthermore reconstruct each of the k microcollisions. Reconstruction is performed by multiplying the frequency domaion representations of our excited modes L and suppressed modes N and upsampling.

4 Attempting A Granular Approach

Based on the two analysis in Figure 1, an alternative approach was attempted which combines certain aspects of both the Lee et al. and the Lagrange et al. models with parameterized granular synthesis. Using the aforementioned microcollisions' extracted excitation signal and ESPRIT approximated filter parameters, a polyphonic granular synthesizer was implemented using Max MSP which stochastically triggers micro-excitations and passes them through second-order biquad sections, simulating modal reconstruction.

The grain generator has a number of parameters that attempt to add variety to the microcollision's excitation. Grains are driven using phase-deviated phasors, allowing microcollisions to polyphonically overlap with each other using Max's mc objects. These grains are passed through a Meixner-esque AD envelope, whose attack, decay, and amplitude are sampled at random. The grains are then passed through peak/notch biquad section polyphonically, each grain periodically exciting one of the 20 modes estimated using ESPRIT. The modes are also programmed to deviate in frequency to simmulate gradual changes in resonnace overtime, although this did not seem to have a huge perceptual impact on the sound. After being passed through the biquad sections, the signal is mixed down and smoothened out slightly.

While the sounds produced by the grain generator were perceived to have the general timbral qualities of both the original marble rolling sound and coin scraping sound, the end result seemed to be a bit too periodic to be believable. Further work must be considered in parameterizing ways in which the excitation grains develop over time. Some statistical procedures that relate to this issue will be discussed in the next section, along with a proposal to map this model to a controller interface.

5 Interpolating and Mapping The Object-Agnostic Approach To An Interactive Scheme

The Lagrange et al. paper touches on the possibility of designing an encoding scheme for generating and convolving dirac impulses as microcollisions, however it is only briefly mentioned. In this section, an interactive scheme is proposed for reproducing sustained contact by using a touch screen tablet as an interactive interface. This approach assumes that the material properties of the surface in question are roughly uniform.

The use of tablet devices and touch screens for musical iteraction can potentially span a wide range of parameters, including multi-touch and basic gestural recognition. This is seen in Serafin et al [9]. in which a Wacom tablet is used to simulate bowing action on a violin. Tablet usage has the potential to map very gracefully to object contact models, since the literal action of interacting with a touch screen is a form of object contact in itself.

The interactive scheme discussed in this section has yet to be implemented, but remains a worthy idea for future work. In the proposed schema, we assume that the interface in question is a basic single-touch tablet with two dimensional coordinates, pressure sensitivity, and acceleration sensing capabilities.

First, in order to improve the object agnostic model based on the findings from Lee et al., we extend our set $\mathbb{T}$ from Section 2 to include dynamically changing modal parameters. In order to find these modal parameters, a similar onset detection and collision extraction step must be taken. This will result in K microcollisions whose modal parameters will be represented by $S_k(z)$, found by performing ESPRIT. The resulting set of filter parameters can be appended to our previous set.

$$\mathbb{T} = \{t_m, a_m, z_{k,m}\} \tag{13}$$

In order to acheive the generation of microcollisions based on touch-screen interaction, encoded variables t_m and a_m must no longer be thought of as time-dependent and instead must be mapped to the parameters of the interface. Findings from the Lagrange et al. paper model both of these variables as statistical distributions based on the time-domain data analysis. Sampling microcollision envelope lengths t_m from the gamma distribution and dirac amplitude values a_m from an exponential distribution were observed to be perceptually sound statistical metrics.

Mapping these distributions to the proposed interface can be done by inversely attenuating the envelope-generating gamma distribution with the interfaces acceleration parameter $\alpha(t)$ and the amplitude-generating exponential distribution with the pressure parameter $\phi(t)$. This schema will thus provide an interaction in which the excitation signal is directly dependent on treating the finger or stylus as an excitation object. Sampled values of t_m and a_m are displayed in (14) and (16) along with the pressure and acceleration attenuations and ranges, while their corresponding cummulative distribution functions (CDF) are shown in (15) and (17).

$$t_m \sim (1 - \alpha(t))\Gamma(k, \theta),\ 0 \leq \alpha(t) \leq 1 \tag{14}$$

$$CDF(x \mid k, \theta) = \int_0^x f(u|k, \theta)du = \frac{\gamma\left(k, \frac{x}{\theta}\right)}{\Gamma(k)} \tag{15}$$

$$a_m \sim \phi(t) EXP(\lambda),\ 0 \leq \phi(t) \leq 1 \tag{16}$$

$$CDF(x \mid \lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{17}$$

Both the statistical approach and time encoded approach generate microcollisions in a way that assumes a parametric independence from the actual physical surface in question. However, when mapping the modal parameters of the system onto a two dimensional tablet, a direct modeling of the geometry of the resonant surface can be approximated. In other words, the filter parameters $z_{k,m}$ can be captured via multiple recordings of the source object moving along the surface in parallel, and mapped to the axis of the tablet. After an interpolation step, the tablet's axis can resemble a literal mapping of the modal properties of the resonant surface in question.

This is expressed in (18), which represents a vector $\underline{x}$ of filter parameters extracted from N recordings of parallel pathways of the source object across the surface. Here, m represents the discrete time encoded analysis windows from (13), only this time taken from multiple recordings. This vector is mapped to $f(m_1, m_2)$ which represents the 2D coordinates of the touchpad. The mapping of filter parameters from the discrete encoded format on the left to a pseudo-continuous mapping on the right requires interpolation of filter parameters across the surface of the touchpad. Thus, we can use something like bicubic interpolation, a 2D interpolation method commonly for image resolution problems. [10]

$$\underline{x} = \begin{bmatrix} z_{k,m,1} \\ z_{k,m,2} \\ z_{k,m,3} \\ \vdots \\ z_{k,m,N} \end{bmatrix}_{m \in M} \rightarrow f(m_1, m_2)_{m_1 m_2 \in \mathbb{R}^2} = \sum_{i=0}^{M} \sum_{j=0}^{N} a_{ij} m_1^i m_2^j \tag{18}$$

6 Conclusion

This paper has focused mainly on two approaches using source-filter techniques in order to model contact between rigid objects. In the prevoius section, these parameters are mapped to a hypothetical interface using the sampling of excitations from statistical distrubutions and the mapping of modes from a time-domain representation to a spatial representation using interpolation and approximation techniques. Despite the fact that these two models adhear to many of the physical variables of object interaction, they are nonetheless statistical models that depend first and foremost on data collection in order to produce a perceptually convincing result.

A different family of approaches for solving the problem of modeling object contact uses a more fine-tuned approach, usually focusing on a specific class of object interaction, often attempting to characterize the physical system using a physics-based continuous time models. These continuous time model are then discretized and evaluated, resulting in sound generation that is exemplary of many physical variables taken into account. This approach takes after a number of successful physical models of acoustic instruments in which physics-based differential equations are discretized in the form of digital waveguides or finite difference methods [5]. This approach can be seen in [11] in which a physics-based simulation of a rolling ball is implemented alongside a corresponding visual and haptic simulation. More recently in [12], two respective models are developed for rolling sounds and scraping sounds, both of implement complex physical models taking into account variables such as normal force, contact force, surface depth maps, and vertical trajectory.

Future work in mapping object interaction models onto electronic musical interfaces might benifit greatly from attempting to map one of the two models proposed in [12] as they are both the most contemporary writing on synthesizing object contact and present incredibly detailed systems that represent many of the physical variables that the source filter models lack.

References

[1] G. S. M. Lagrange and P. Depalle, "Analysis/synthesis of sounds generated by sustained contact between rigid objects," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 509–518, March 2010.

[2] P. D. J. Lee and G. Scavone, "Analysis / synthesis of rolling sounds using a source-filter approach," *13th International Conference on Digital Audio Effects, DAFx 2010 Proceedings*, vol. 18, January 2010.

[3] B. D. N. B. J. E. O. D. S. M. M. Lagrange, R. Badeau and L. Daudet, "The desam toolbox: Spectral analysis of musical audio," *13th International Conference on Digital Audio Effects, DAFx 2010 Proceedings*, October 2010.

[4] K. van den Doel and D. K. Pai, "Modal synthesis for vibrating objects," 2007.

[5] J. O. Smith, *Physical Audio Signal Processing.* W3K Publishing, 2010.

[6] R. B. R. Badeau and B. David, "Eds parametric modeling and tracking of audio signals," *Proc. of the 5th International Conference on Digital Audio Effects (DAFx)*, pp. 139–144, January 2002.

[7] M. V. Ronkin, A. A. Kalmykov, S. O. Polyakov, and V. S. Nagovicin, "Numerical analysis of adaptive signal decomposition methods applied for ultrasonic gas flowmeters," in *AIP Conference Proceedings*, vol. 2425, no. 1. AIP Publishing LLC, 2022, p. 130009.

[8] G. Scavone, *modesynth.m*, ver. 2, 2004-2022[Online]. [Online]. Available: http://www.music.mcgill.ca/ gary/307/matlab/modesynth.m

[9] S. Seraan, R. Dudas, M. Wanderley, and X. Rodet, "Gestural control of a real-time physical model of a bowed string instrument," 11 1999.

[10] P. Breeuwsma, "Cubic interpolation." [Online]. Available: https://www.paulinternet.nl/?page=bicubic

[11] M. Rath and D. Rocchesso, "Continuous sonic feedback from a rolling ball," *IEEE MultiMedia*, vol. 12, pp. 60–69, April 2005.

[12] J. T. V. Agarwal, M. Cusimano and J. McDermott, "Object-based synthesis of scraping and rolling sounds based on non-linear physical constraints," *24th International Conference on Digital Audio Effects (DAFx)*, vol. 18, pp. 136–143, January 2021.