OBJECT-BASED SYNTHESIS OF SCRAPING AND ROLLING SOUNDS BASED ON NON-LINEAR PHYSICAL CONSTRAINTS

Vinayak Agarwal *

Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, US
`vinayaka@mit.edu`

Maddie Cusimano

Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, US
`mcusi@mit.edu`

James Traer

Department of Psychological and Brain Sciences
University of Iowa
Iowa City, US
`james-traer@uiowa.edu`

Josh McDermott

Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, US
`jhm@mit.edu`

ABSTRACT

Sustained contact interactions like scraping and rolling produce a wide variety of sounds. Previous studies have explored ways to synthesize these sounds efficiently and intuitively but could not fully mimic the rich structure of real instances of these sounds. We present a novel source-filter model for realistic synthesis of scraping and rolling sounds with physically and perceptually relevant controllable parameters constrained by principles of mechanics. Key features of our model include non-linearities to constrain the contact force, naturalistic normal force variation for different motions, and a method for morphing impulse responses within a material to achieve location-dependence. Perceptual experiments show that the presented model is able to synthesize realistic scraping and rolling sounds while conveying physical information similar to that in recorded sounds.

1. INTRODUCTION

Collisions, scraping, and rolling are commonplace in daily life, and the sounds they produce convey information about physical events in the world. Often this information is uniquely available via sound. For instance, the visual evidence that an object is in contact with another is often ambiguous, but objects in contact will make sound if they move.

The synthesis of contact sounds is accordingly important for a wide range of applications including virtual reality, game engines, auditory displays, and the training of machine perception systems. Such synthesis requires mapping physical variables (object materials, shapes, and motions) to sound. These generative models of sound could also contribute to theories of perception, which involves inferring physical causes from sound [1], potentially by inverting internal generative models [2].

Contact sound synthesis is possible via detailed but expensive physical simulations [3, 4, 5, 6, 7, 8, 9]. However, many applications require more efficient synthesis, as might be enabled by lower-dimensional characterizations of objects. Previous methods have either used signal-based heuristics [10, 11] or physics-inspired intuitive modelling [12, 13, 14, 15]. These models are able to convey some intended physical parameters, but thus far have not been fully perceptually convincing.

We previously introduced a method to synthesize impact sounds by characterizing material-specific impulse response distributions, sampling from them, and convolving them with simple spring-based models of impact forces [16]. In that paper, we extended the method to synthesize scraping sounds using force derived from a surface depth map, but the sounds produced were not fully realistic, sounding excessively rough and not producing a fully accurate sense of motion.

Here we introduce a new method to synthesize scraping and rolling sounds from object properties. The key innovations are a nonlinearity in the calculation of the contact force from the surface depth map and the use of normal forces that more accurately reflect typical scraping motions. These are combined with a method for generating more realistic location-dependent impulse responses via morphing within the same material, similar to that proposed in [17], and a periodic force produced by rolling objects due to misalignment of the center-of-mass with the geometric center [18, 10]. We term the approach 'object-based', as it relies on perceptually-relevant macroscopic properties of objects and their motions.

The resulting sounds are substantially more realistic and recognizable than those from previous methods. We first describe and evaluate our scraping synthesis. We then show that the scraping synthesis can be extended to generate realistic rolling sounds. You can listen to demos of sounds from all our experiments at `http://mcdermottlab.mit.edu/scraping_rolling.html`.

2. SCRAPING SOUND SYNTHESIS MODEL

We use a source-filter approach depicted in Figure 1. Similar to [16, 11, 19], we generate a sound waveform by the convolution of a contact force function and a net impulse response (IR). The contact force $f(t)$ is obtained by modelling the microscopic trajectory of the scraper on the scraped surface. We use a weighted sum of the impulse responses of both the scraper and the scraped surface to get the net impulse response of the interaction. To reflect the variation in object resonances with position, the impulse
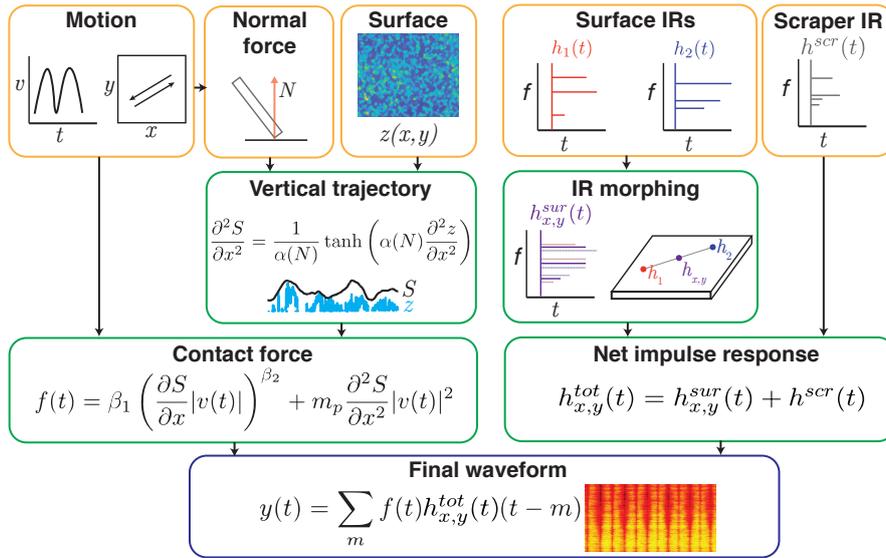
Figure 1: *Complete scraping synthesis model. Yellow boxes: inputs to model. Green boxes: intermediate representations computed from inputs. Blue box: sound waveform computed from contact force and net impulse response.*

response of the surface changes continuously over the motion. We describe the force, impulse response and audio waveform calculation in more detail below.

2.1. Contact force for scraping

The contact between scraper and surface causes the two bodies to undergo continuous micro-collisions. Previous authors have sampled the friction force as $1/f^\beta$ noise [19, 10] or from a distribution over concentrated impact events [11, 20, 21]. In our model, the only major assumption is that the scraper follows the surface without bouncing. We also assume that the masses of the scraper and surface do not change over the course of the motion.

The microscopic interaction force is modelled in two parts - a vertical force f_v (from pressing the scraper down) and a horizontal force f_h (from horizontal micro-collisions with surface asperities). We also calculate an additional macroscopic normal force (§2.3), which modulates the microscopic interaction force.

The vertical force f_v is estimated as:

$$f_v(t) = m_p \ddot{S}(x, y) \tag{1}$$

$$= m_p \left(\frac{\partial^2 S(x, y)}{\partial x^2} |v_x(t)|^2 + \frac{\partial^2 S(x, y)}{\partial y^2} |v_y(t)|^2 \right) \tag{2}$$

where m_p is the mass of the scraper, $S(x, y)$ is the scraper's vertical trajectory over the surface, and $v_x(t)$ and $v_y(t)$ are the velocities of the scraper in the x and y directions respectively (which define the plane of the surface).

We assume the horizontal contact force f_h to originate from the horizontal collisions of the scraper and steep asperities of the surface texture. Based on previous literature [16], we assume the following relation:

$$f_h(t) = \beta_1 \left| v_x(t) \frac{\partial S(x, y)}{\partial x} + v_y(t) \frac{\partial S(x, y)}{\partial y} \right|^{\beta_2} \tag{3}$$

where the partial derivatives with respect to the position (x, y) reflect the slopes of the surface texture and determine the intensity of

the micro-impacts. The velocities v_x and v_y determine the density of these impacts in time. β_1, β_2 are free parameters that we set to 0.05 and 1, respectively.

We model the contribution of the horizontal and vertical force components to the resulting sound as a convolution with the same impulse response. Because the convolution operator is linear and the resulting sound waveforms add together ($y(t) = f_v(t) * h(t) + f_h(t) * h(t)$), for notational convenience we denote a total force term given by a scalar sum $f_{tot}(t) = f_v + f_h$.

2.2. Trajectory of the scraping object

One of the most critical components of our model is the vertical trajectory $S(x, y)$ of the scraper near the surface. Most previous models have assumed the scraper's vertical trajectory to be the same as the surface depth profile $z(x, y)$ [16, 11]. In our model, we use $z(x, y)$ to derive the estimated vertical trajectory of the scraper but additionally incorporate nonlinear physical constraints on the scraper-surface interaction. As in previous work, we assume the scraper to be in contact with the surface at a single point.

We obtained surface depth profiles for several everyday materials using a scanning confocal microscope (Keyence VK-X260K, horizontal resolution of 5.6 μm, vertical resolution of 0.1 nm) [16]. In the future, we plan to synthesize surface depth maps as textures using relevant texture statistics of measured depth profiles [22].

The assumption that the scraper exactly follows the surface depth profile z is physically unrealistic for two reasons: 1) the forces involved would be exceptionally high near the locations on the surface where the slopes change rapidly, and 2) both the surface and scraper materials would have to be abnormally elastic.

To solve these issues, we have to constrain the allowed curvatures of the scraper trajectory $S(x, y)$ to limits that are physically realizable given bounded applied force. The applied force bounds the curvature of the turns that the scraper can take over the surface asperities, limiting its acceleration and hence the second derivative of S (Figure 2). To achieve this, we used a tanh non-linearity with
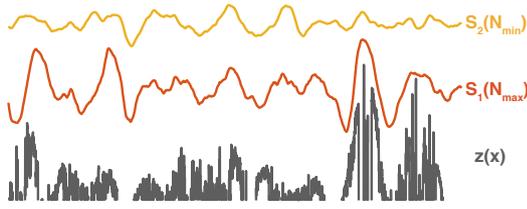
Figure 2: *Trajectory S of the scraper point. The trajectory is determined by the surface depth profile ($z(x)$, shown in grey) and the normal force N. Larger normal forces (red) produce more extreme scraper trajectories than smaller normal forces (yellow).*

limits dictated by the elasticity of the material and the macroscopic normal force N.

$$\frac{\partial^2 S(x,y)}{\partial x^2} = \frac{1}{\alpha_x(N)} \tanh\left(\alpha_x(N)\frac{\partial^2 z(x,y)}{\partial x^2}\right) \quad (4)$$

$$\frac{\partial^2 S(x,y)}{\partial y^2} = \frac{1}{\alpha_y(N)} \tanh\left(\alpha_y(N)\frac{\partial^2 z(x,y)}{\partial y^2}\right) \quad (5)$$

α_x and α_y change the limits of the non-linearity depending on the normal force:

$$\alpha(N) = (1-\nu)\alpha_{\max} + \nu\alpha_{\min} \quad (6)$$

$$\nu = \left(\frac{N - N_{\min}}{N_{\max} - N_{\min}}\right)^\zeta \quad (7)$$

$N_{\max}$ and $N_{\min}$ are determined by the variation in normal force described in the next section; for constant N we substitute fixed values for α_x and α_y. The current synthesis process is dependent on a judicious choice of ζ, $\alpha_{\max}$ and $\alpha_{\min}$ (we found values of 0.95, 0.05, and 0.01, respectively, to yield realistic sounds). In the future, we aim to estimate these parameters from the applied vertical force and hardness of the material.

This procedure is a form of soft clipping which introduces high-frequency components, presumably because the tanh function does not limit the higher-order derivatives of the scraper trajectory to an extent that is physically correct. To obtain realistic sounds, we found it necessary to apply a Gaussian moving average following the non-linearity to remove these high-frequency components. The window size in each dimension was proportional to α_x and α_y, with an average half window size of 5 samples at a sampling rate of 44.1kHz. Alternative choices of non-linearity might obviate the need for this additional step.

2.3. Effect of macroscopic normal force

The applied force dictates how closely the scraper tracks the surface, which in our model is reflected in the force-dependent non-linearity parameters (α_x, α_y) that influence the modulation of the normal force (Figure 2). The applied force can vary over the scraping trajectory, in part due to constraints on how the human hand applies forces. For example, consider back-and-forth scraping with a cylindrical object held by hand (Figure 3). This action can be approximated as simple harmonic motion with an angular frequency ω. We calculate the resulting normal force assuming that 1) the applied force is along the length of the cylinder 2) the force is proportional to the horizontal acceleration of the cylinder, and 3) the cylinder is at an angle θ from the horizontal. As a result, the

force is maximal when close to the torso and minimal when furthest away. The normal forces at both ends of the trajectory are then given by:

$$N_{\max} = \frac{mg + \omega^2 mL\tan\theta}{1 - \mu\tan\theta} \quad (8)$$

$$N_{\min} = \frac{mg - \omega^2 mL\tan\theta}{1 - \mu\tan\theta} \quad (9)$$

This normal variation is characteristic of back-and-forth scraping motion (e.g. from sanding or scrubbing). The normal force variation influences the scraping sound via α_x and α_y as described in the previous section. For each motion trajectory, we similarly modelled the normal force variation assuming a scraping cylinder held at a fixed angle to the surface.
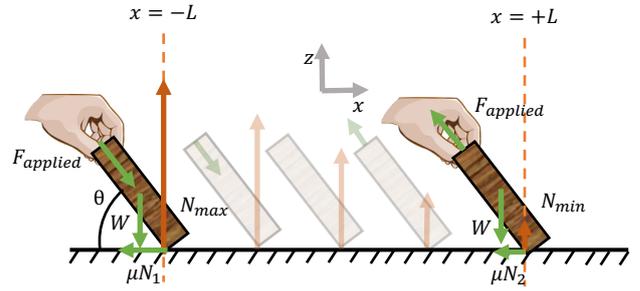
Figure 3: *Macroscopic forces on the scraper when undergoing human-produced simple harmonic motion (as when scraping a held object back and forth on a surface). The normal force (red vector) varies with position due to the variation in the applied force $F_{applied}$.*

2.4. Morphing impulse responses

The impulse responses on the surface depend on the location on the surface currently being excited, because the boundary conditions change with the scraper position. Our measurements of object impulse responses indicated that the modal frequencies and their powers typically vary smoothly with location. To generate an impulse response for an arbitrary surface location, we thus morph from one measured impulse response to another by smoothly interpolating between both the modal frequencies and powers (similar to [17]), in contrast to our previous work where we cross-faded impulse responses [16].

We used impulse responses measured in previous work [16] at different locations on the surfaces of objects. We then extracted the mode parameters of the 50 strongest modes using sinusoidal modelling, giving us the frequency f_i of each mode and its time-dependent mode power $A_i(t)$. To obtain the mode parameters for an impulse response at a location intermediate between two measured impulse responses we logarithmically interpolated between the measured modes in both frequency and amplitude. The impulse response for the location was then synthesized as a superposition of decaying sinusoids using these morphed parameters.

As the scraper applies a force on the surface, the surface applies an equal and opposite reaction force on the scraper. If the scraper is not damped, it also contributes to the overall sound because of the effect of this forcing on the scraper's resonances. To account for this, we sum the synthetic impulse response calculated

using morphed parameters with the measured impulse response of the scraping object weighted by η, which determines the relative amplitude of the surface and scraper impulse responses.

Since each position on the surface has a different impulse response, and because the scraper moves continuously, the traditional linear view of the sound waveform being the result of a convolution of force and a fixed impulse response does not hold. To accommodate a different impulse response at each location, we calculate the waveform $y(t)$ using a modified convolution:

$$h_{x,y}^{\text{sur}}(\tau) = \begin{cases} \sum_{i=1}^{50} A_i(\tau) \sin(2\pi f_i \tau) & 0 \le \tau \le t_0 \\ 0 & otherwise \end{cases} \quad (10)$$

$$h_{x,y}^{\text{tot}}(\tau) = h_{x,y}^{\text{sur}}(\tau) + h^{\text{scr}}(\tau) \quad (11)$$

$$f(t) = \begin{cases} f_{tot}(t) & 0 \le t \le t_1 \\ 0 & otherwise \end{cases} \quad (12)$$

$$y(t) = \begin{cases} \sum_{m=1}^{t} h_{x,y}^{\text{tot}}(t-m)f(m) & 0 \le t \le t_0 \\ \sum_{m=1}^{t_0} h_{x,y}^{\text{tot}}(t_0-m)f(t+m-t_0) & t_0 \le t \le t_1 \\ \sum_{m=1}^{t_2-t} h_{x,y}^{\text{tot}}(t_0-m)f(t+m-t_0) & t_1 \le t \le t_2 \end{cases} \quad (13)$$

where $t_2 = t_0 + t_1$, $h^{\text{scr}}(t)$ has the same form as $h_{x,y}^{\text{sur}}(t)$ but is independent of position, $h_{x,y}^{\text{tot}}(t)$ is the total impulse response when the scraper is at (x,y), and f_{total} is the excitation at the surface.

3. PERCEPTION OF SYNTHETIC SCRAPING

To assess whether our synthesis model produces perceptually compelling scraping sounds, we conducted two psychophysical experiments in which listeners made judgments about synthesized scraping sounds. We compared our synthesis model to several 'ablated' versions of the model, in which we omitted model components in order to test their perceptual relevance. For baselines, we also compared our model to our implementations of previous synthesis methods based on (1) physical synthesis [16], (2) signal manipulation [11], and (3) minimalist modulated noise that has previously been shown to elicit accurate motion judgments [23]. These experiments were conducted online using Amazon's Mechanical Turk platform, using a standardized test to verify that participants were wearing headphones [24]. We have previously found that online participants can perform about as well as in-laboratory participants [25, 26, 27] provided basic steps are taken to maximize the chances of reasonable sound presentation by testing for earphone/headphone use and to ensure compliance with instructions. All participants had self-reported normal hearing and had IP addresses from North America.

3.1. Experiment 1. Realism of synthetic scraping sounds

We first tested whether our model could convincingly render descriptions of physical scraping events, by asking listeners to rate the realism of synthetic scrapes. Participants were presented with a text description of the scraping event, which specified the material of the scraper, its motion, and the material of the surface (Fig. 4A, top). The scraper material was either PVC or poplar

wood and the surface material was basswood, poplar wood, or ceramic. The motion was one of five variants: slow back-and-forth, fast back-and-forth, circular, short single scrape, and long single scrape. Each listener evaluated a total of 30 trials comprising a fully-crossed set of these parameters (2 scraper materials × 3 surface materials × 5 motions). For each description, listeners rated the realism of seven synthetic sounds, each generated with a different method. The methods were (1) the full model, (2) an ablated model using only the surface impulse response ($\eta = 0$), (3) an ablated model lacking variation in the normal force (omitting §2.3) (4) an ablated model without a non-linearity to constrain the trajectory curvatures (omitting §2.2), (5-7) the baseline models. All seven sounds were presented in a single trial using a MUSHRA-like paradigm (Fig. 4A). The order of the sounds was randomized on each trial. We did not include audio recordings in this experiment because they typically have distinguishing features that do not relate to scraping (e.g. room noise).

The results (Fig. 4B) show that our model produces scraping sounds that are more realistic than previous scraping synthesis methods. The non-linearity ablation caused the biggest decrease in the realism. The 'Only surface IR' and normal force ablations produced relatively small impairments of the realism, potentially because there exist alternative realistic physical explanations of those sounds (e.g. damping of the scraper).
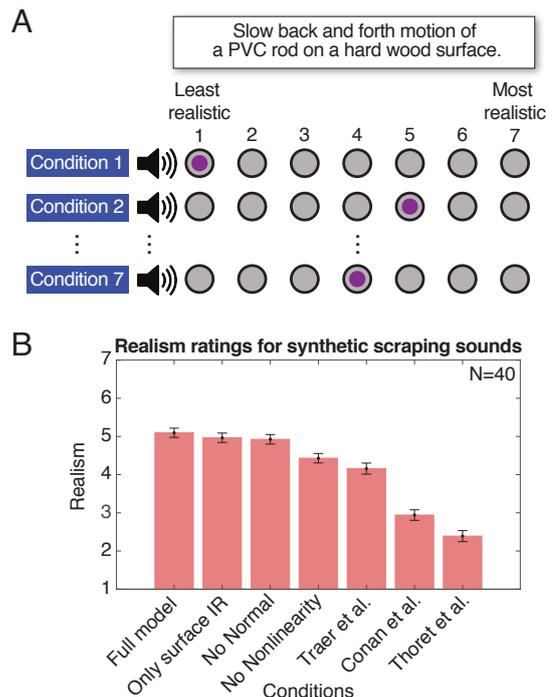
Figure 4: *Experiment 1: Realism of synthesized scraping. A) Participants (N=40; mean age = 38.3 years) rated the realism of 7 different renderings of an object scraped back-and-forth over a surface, using a MUSHRA paradigm. Each rendering was via a different synthesis method. Participants were given a text description of the scraping event. B) Results of Experiment 1, showing mean realism for each synthesis method. Error bars plot SEM.*

3.2. Experiment 2. Perception of motion from scraping sounds

We next asked whether participants could infer the motion that produced the scrape. In addition to testing whether the synthesis conveys an important physical parameter, this experiment provides a more fine-grained examination of the effect of the various ablations as it allows analysis of confusions between different motions.

Listeners were presented with a single scraping sound and then asked to choose the most probable path (Fig. 5A) traced by the object from five options: 1) a scribble, 2) four scrapes in a straight line, 3) four back-and-forth scrapes, 4) a single long line, or 5) a circle (Fig. 5B). The scribble motion was obtained by optical tracking [16] and the other motions were modelled using ideal trajectories. We chose these motions because they provide pairs which share a velocity profile, but differ spatially: back-and-forth and four-in-one line, and the circle and single line. Based on previous literature [23] we expected that listeners would make mistakes within these pairs but would make accurate judgments between them. We tested the seven synthesis conditions of Experiment 1 as well as real audio recordings.

As expected, the confusion matrices (Fig. 5C) show that listeners experienced confusions when listening to recorded as well as synthesized audio. However, the confusion patterns for our full model are more similar to those of recorded audio than were the other synthesis methods. We quantified this similarity between the confusion matrix for recorded audio $C_{recorded}$ and each synthesis method $C_{synth,i}$ as follows:

$$Similarity(i) = 1 - \frac{\|C_{recorded} - C_{synth,i}\|_2}{\|C_{recorded}\|_2} \quad (14)$$

As shown in Figure 5D, the full model achieves the highest similarity to listeners' judgments on recorded audio. The 'Only surface IR' model performs similarly to the full model, while the normal force and non-linearity ablations greatly decrease the similarity. These results indicate that capturing naturalistic variation in the normal force and constraining the trajectory curvature are important for the perceptual estimation of motion from scraping.

4. ROLLING SOUND SYNTHESIS MODEL

To synthesize rolling sounds, we use a similar source-filter model where the filter changes with time due to the change in the position of the rolling object, but with a contact force that differs from that for scraping. Pure rolling is a unique form of sustained contact where the surface is excited without the point of contact on the object being in relative motion with the surface. Previous efforts to synthesize rolling sounds have proposed a force term arising from the offset of the object's centre of mass from its geometric center (due to slight deviations from perfect sphericity). We hypothesized that there would be an additional contribution from a scraping-like term due to the catching and release between surface asperities that would depend on the trajectory of the point in contact. Figure 6 depicts a schematic of the overall contact force for rolling. The first two terms reflect the same calculation used in scraping, and the third term is unique to rolling.

4.1. Contact force for rolling

We used a rolling-specific force term similar to that previously proposed [18, 10], including a non-linear dissipative component.

Our implementation of this force differed from prior work in being based on the trajectory of the ball, derived as in our scraping model as a nonlinear function of the surface depth map:

$$f_{roll}(\rho, \dot{\rho}) = k\rho^{3/2} + \lambda\rho^{3/2}\dot{\rho} \quad (15)$$

where ρ denotes penetration depth (the deviation of the ball center of mass from its mean vertical position), k denotes the equivalent spring constant of the ball material and λ is a dissipation constant which we set by ear. ρ and its derivative are given by

$$\rho = R - r\cos\frac{x}{R} + S(x, y) \quad (16)$$

$$\dot{\rho} = \frac{r}{R}\dot{x}\sin\frac{x}{R} + \dot{x}\frac{\partial S(x, y)}{\partial x} \quad (17)$$

where x is the horizontal position of the center of mass:

$$x = R\theta - r\sin\theta \quad (18)$$

with r the distance between the center of mass and the geometric center, R the mean radius of the ball, and θ the angular displacement of the ball.

The total contact force is given by

$$f(t) = \beta_1\left|\frac{\partial S}{\partial x}v(t)\right|^{\beta_2} + m_p\frac{\partial^2 S}{\partial x^2}|v(t)|^2 + (k\rho^{3/2} + \lambda\rho^{3/2}\dot{\rho}) \quad (19)$$

The relative contribution of the scraping and rolling terms likely depends on how much slip is present in the rolling motion and the roughness of the materials, and can be adjusted using the existing free parameters β_1, β_2, k and λ .

For all three force terms, the vertical trajectory $S(x, y)$ of the point of contact on the surface is determined as in the scraping model. A tanh non-linearity was used to constrain the curvatures of the vertical trajectory based on the amount of normal force. The only difference is that the normal force varies periodically over the rolling interaction due to the offset between the center of mass and the geometric center of the sphere, being maximal when the centre of mass moves downwards, and minimal when it moves up (Figure 6). This variation in the normal force affects the penetration of the ball into the surface, which we modeled by varying the non-linearity parameter α and the subsequent smoothing.

4.2. Impulse responses

Location-dependent impulse responses for rolling were synthesized in the same way as for scraping. An impulse response for the rolling object is added to the surface impulse response, with the relative weighting depending on the ball material. At present we set the relative weighting by hand. In the rolling sound recordings we sought to emulate in our experiments, the surfaces were planks, and typically had a much higher contribution to the overall sound than the balls. If modeling the sound of a ball rolling on a floor, which is typically damped, the weighting would instead upweight the contribution of the ball.

5. PERCEPTION OF SYNTHETIC ROLLING

To evaluate the synthesis model, we asked human listeners to rate the realism of synthetic sounds generated in various ways, using the MUSHRA paradigm from Experiment 1. We again compared our model to ablated models with omitted components and to two previous baselines using types of physics-based [18] and signal-based synthesis [11].
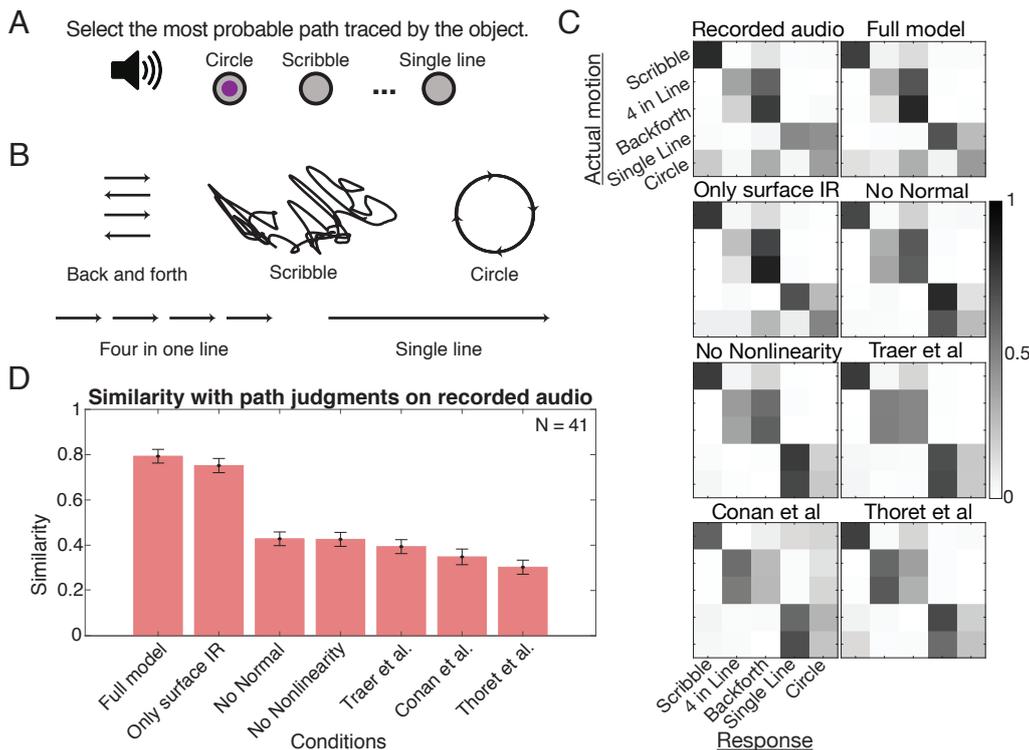
Figure 5: *Experiment 2: Motion recognition from scraping. A) Participants (N=41; mean age = 40.3 years) listened to a sound and selected the path traced by the scraping object. B) Sounds were generated for each of five possible motions. C) Confusion matrices for each of the eight conditions. D) Similarity of the confusion matrix of each synthesized condition compared to that for the recorded sounds.*

5.1. Experiment 3. Realism of synthetic rolling sounds

In each trial, online participants were presented with a test description of a rolling event which specified the material of the rolling object and the surface, incline of the surface and the motion of the object (Figure 7A). The ball material was ceramic, glass or wood. The surface material ceramic, bass wood or poplar wood. The surface had a gradual incline in all the sounds and the ball was rolled up or down the incline. The different synthesis conditions were (1) full model, ablated models with (2) only the surface IR contribution, (3) a constant surface IR (omitting the location-dependent morphing), (4) only the ball IR contribution, (5) no nonlinearity in the calculation of contact force from the depth map, and (6-7) baseline methods [18, 11].

As shown in Figure 7B, our model produced more realistic rolling sounds than the previous baseline models. The surface resonance was more important than the ball resonance for overall realism ('Only Surface IR' vs. 'Only Ball IR'). As with the scraping sounds of Experiment 1, the non-linearity had a significant impact on the realism of rolling sounds (without it, sounds were unrealistically rough). In contrast, morphing the impulse responses did not significantly contribute to the realism.

6. DISCUSSION

We developed a novel method for synthesizing scraping and rolling sounds and evaluated it with perceptual experiments. The model is object-based, in the sense of depending on relatively macroscopic

properties of objects and their motions. The key innovations compared to previous methods are the introduction of a nonlinearity in the relationship between contact force and the surface depth map and the use of normal forces that more accurately reflect scraping by biological organisms. We combine these two ideas with variants of several previous proposals for synthesis in this domain. The resulting synthesis is substantially more realistic than previous methods.

Our experiments explored the contribution of different components of the synthesis model. We found that the non-linearity used to constrain the curvatures of the vertical trajectory of the scraper (and thus the resulting contact force) was critical. Without the nonlinearity, the contact force is unrealistically high and the sounds are unrealistically rough.

The normal force variation was also important for compelling synthesis. Its effect was most evident in the motion estimation experiment, where it was necessary to convey motion comparable to actual audio recordings. Motion recognition for the full model containing this normal force variation qualitatively matched that for audio recordings, despite using ideal velocity profiles which do not follow empirical relations for human-generated profiles [23].

Several synthesis components did not produce a clear benefit in some of the experiments, but this may be be a limitation of the experiments rather than of the synthesis. For instance, realism judgments for scraping did not distinguish between the full model and the ablated model without the normal force variation. This is plausibly because sounds without variation in the normal force correspond to an alternative physically possible situation in
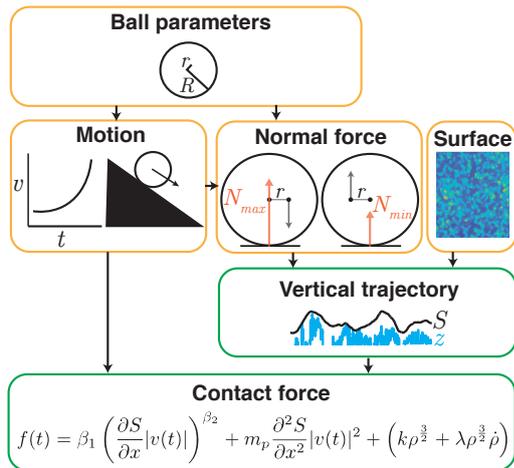
Figure 6: *Synthesis of contact force for rolling. Yellow boxes are inputs to the synthesizer, which are combined to yield the contact force. The contact force is combined with location-specific impulse responses in the same manner as for scraping (shown in Figure 1).*

Figure 7: *Experiment 3: Realism of synthesized rolling. A) Participants (N=33; mean age = 36.2 years) rated the realism of 7 different renderings of a ball rolling down or up an inclined surface, using a MUSHRA paradigm. Each of the 7 renderings was from a different synthesis method. The ball and surface varied in material. B) Results of Experiment 3, showing mean realism for each synthesis method. Error bars plot SEM.*

which the normal force does not change over the path. Participants would rate the sound as realistic because they can envision a physical scenario that could have produced the sound, even though that scenario deviates from the one that was intended. The same issue may explain why there was little effect of omitting the contribution of the scraper impulse response, as this could correspond to a situation with damping.

We suspect that the failure to demonstrate a benefit of location-dependent impulse response morphing (Experiment 3; Fig. 7) is also a limitation of the realism judgments we used. There are realistic physical explanations for both the cases where impulse responses change with position and where they do not. For example, rolling on a damped floor would not lead to location-dependent impulse responses, whereas rolling on a small wooden plank yields a much more appreciable location-dependence. Overall, our subjective sense is that all components of the synthesis that we tested here contribute meaningfully to the resulting sound, but additional work is needed to better understand and demonstrate the situations in which each component of the synthesis is most perceptually relevant.

The model presented here can be further improved in several respects. At present several parameters of the model are set by hand. For instance, we currently lack an empirical relationship between the non-linearity parameter and the normal force. We plan to investigate the impact of the softness/hardness of the materials empirically and to base this parameter on empirical measurements in the future. At present we also set the balance between the rolling and scraping components of rolling sounds by hand, when this should ultimately be determined by physical parameters of the motion and surface characteristics. It also seems likely that the perception of scraping and rolling is based on summary statistics [28] that capture surface roughness (as opposed to detailed representations of the underlying depth maps). Synthesis could thus plausibly be based on depth maps synthesized from summary statistics, which would make for a more parsimonious model.

Efficient and high-fidelity synthesis will open the door to new perceptual studies of ecological audition by enabling sound gen-

eration for physical events whose properties can be varied in a controlled manner. Relatively simple synthesis methods with input variables that are both physically meaningful and perceptually consequential can also provide the foundation for models of human auditory perception. If prior distributions are determined for the physical variables, the principles of Bayesian inference can be leveraged to invert the synthesis method and thereby estimate latent physical variables from sound [29, 2]. We plan to develop a complete contact sound synthesis model (including scraping, rolling, and impacts) within an inference framework that can be used to make predictions about the mechanisms underlying human physical inference. In principle, this could be combined with inverse graphics [30, 31] to achieve multimodal physical event perception. Bayesian inference could also provide a way to set the values of the free parameters, given corpora of recorded contact sounds.
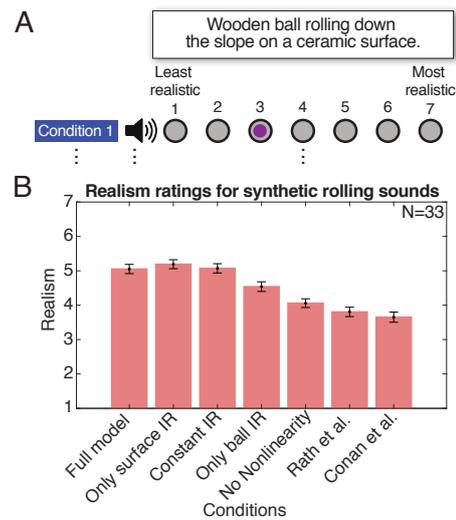
7. REFERENCES

[1] William W Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[2] Daniel Kersten and Alan Yuille, "Bayesian models of object perception," *Current Opinion in Neurobiology*, vol. 13, pp. 1–9.

[3] Claude Cadoz, Annie Luciani, and Jean Loup Florens, "CORDIS-ANIMA: a modeling and simulation system for sound and image synthesis: the general formalism," *Computer music journal*, vol. 17, no. 1, pp. 19–29, 1993.

[4] James F O'Brien, Chen Shen, and Christine M Gatchalian, "Synthesizing sounds from rigid-body simulations," in *Pro-*

ceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation. ACM, 2002, pp. 175–181.

[5] Changxi Zheng and Doug L. James, "Toward high-quality modal contact sound," *ACM Trans. Graph.*, vol. 30, no. 4, July 2011.

[6] Stefan D Bilbao, *Numerical Sound Synthesis*, Wiley Online Library, 2009.

[7] Doug L James, Jernej Barbič, and Dinesh K Pai, "Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources," in *ACM Transactions on Graphics (TOG)*. ACM, 2006, vol. 25, pp. 987–995.

[8] Dinesh Manocha and Ming C Lin, "Interactive sound rendering," in *2009 11th IEEE International Conference on Computer-Aided Design and Computer Graphics*. IEEE, 2009, pp. 19–26.

[9] Nikunj Raghuvanshi and Ming C Lin, "Interactive sound synthesis for large scale environments," in *Proceedings of the 2006 symposium on Interactive 3D graphics and games*. ACM, 2006, pp. 101–108.

[10] Kees Van Den Doel, Paul G Kry, and Dinesh K Pai, "Foleyautomatic: physically-based sound effects for interactive simulation and animation," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 537–544.

[11] Simon Conan, Etienne Thoret, Mitsuko Aramaki, Olivier Derrien, Charles Gondre, Solvi Ystad, and Richard Kronland-Martinet, "Intuitive synthesizer of continuous-interaction sounds: Rubbing, scratching, and rolling," *Computer Music Journal*, vol. 38, no. 4, pp. 24–37, 2014.

[12] Stefania Serafin, *The sound of friction: Real time models, playability and musical applications*, Ph.D. thesis, Department of Music, Stanford University, 2004.

[13] Christophe NJ Stoelinga, *A psychomechanical study of rolling sounds*, Ph.D. thesis, ENSTA ParisTech, 2007.

[14] Federico Avanzini, Stefania Serafin, and Davide Rocchesso, "Interactive simulation of rigid body interaction with friction-induced sound generation," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1073–1081, 2005.

[15] Pedro Sanchez and Joshua Reiss, "Real-time synthesis of sound effects caused by the interaction between two solids," *Journal of the Audio Engineering Society*, March 2019.

[16] James Traer, Maddie Cusimano, and Josh H. McDermott, "A perceptually inspired generative model of rigid-body contact sounds," *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx-19)*, Sep 2019.

[17] Laurent Pruvost, Bertrand Scherrer, Mitsuko Aramaki, Sølvi Ystad, and Richard Kronland-Martinet, "Perception-based interactive sound synthesis of morphing solids' interactions," in *SIGGRAPH Asia 2015 Technical Briefs*. ACM, 2015, p. 17.

[18] M. Rath and D. Rocchesso, "Continuous sonic feedback from a rolling ball," *IEEE MultiMedia*, vol. 12, no. 2, pp. 60–69, 2005.

[19] Zhimin Ren, Hengchin Yeh, and Ming C. Lin, "Synthesizing contact sounds between textured models," in *2010 IEEE Virtual Reality Conference (VR)*, 2010, pp. 139–146.

[20] Mathieu Lagrange, Gary Scavone, and Philippe 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, 2010.

[21] Jung Suk Lee, Philippe Depalle, and Gary Scavone, "Analysis/synthesis of rolling sounds using a source-filter approach," in *13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria*, 2010.

[22] Josh H McDermott and Eero P Simoncelli, "Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis," *Neuron*, vol. 71, no. 5, pp. 926–940, 2011.

[23] Etienne Thoret, Mitsuko Aramaki, Richard Kronland-Martinet, Jean-Luck Velay, and Solvi. Ystad, "From sound to shape: auditory perception of drawing movements.," *Journal of Experimental Psychology: Human Percept ion and Performance*, vol. 40, no. 3, pp. 983–994, 2014.

[24] Kevin JP Woods, Max H Siegel, James Traer, and Josh H McDermott, "Headphone screening to facilitate web-based auditory experiments," *Attention, Perception, & Psychophysics*, vol. 79, no. 7, pp. 2064–2072, 2017.

[25] Kevin J.P. Woods and Josh H. McDermott, "Schema learning for the cocktail party problem," *Proceedings of the National Academy of Sciences*, vol. 115, pp. E3313–E3322, 2018.

[26] Richard McWalter and Josh H. McDermott, "Illusory sound texture reveals multi-second statistical completion in auditory scene analysis," *Nature Communications*, vol. 10, pp. 5096, 2019.

[27] Malinda J. McPherson and Josh H. McDermott, "Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory," *Proceedings of the National Academy of Sciences*, vol. 117, pp. 32169–32180, 2020.

[28] Josh H McDermott, Michael Schemitsch, and Eero P Simoncelli, "Summary statistics in auditory perception," *Nature Neuroscience*, vol. 16, no. 4, pp. 493, 2013.

[29] Maddie Cusimano, Luke B Hewitt, Josh Tenenbaum, and Josh H McDermott, "Auditory scene analysis as bayesian inference in sound source models.," in *Proceedings of the 2018 Meeting of the Cognitive Science Society*, 2018.

[30] Ilker Yildirim, Mario Belledonne, Winrich Freiwald, and Josh Tenenbaum, "Efficient inverse graphics in biological face processing," *Science advances*, vol. 6, no. 10, pp. eaax5979, 2020.

[31] Chuang Gan, Jeremy Schwartz, Seth Alter, Martin Schrimpf, James Traer, Julian De Freitas, Jonas Kubilius, Abhishek Bhandwaldar, Nick Haber, Megumi Sano, Kuno Kim, Elias Wang, Damian Mrowca, Michael Lingelbach, Aidan Curtis, Kevin Feigelis, Daniel M. Bear, Dan Gutfreund, David Cox, James J. DiCarlo, Josh McDermott, Joshua B. Tenenbaum, and Daniel L. K. Yamins, "Threedworld: A platform for interactive multi-modal physical simulation," *arXiv preprint arXiv:2007.04954*, 2020.