Intuitive Physical Inference from Sound

James Traer (jtraer@mit.edu)
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave.
Cambridge, MA 02139 USA

Josh H McDermott (jhm@mit.edu)
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave.
Cambridge, MA 02139 USA

Abstract:

When objects collide they produce sound. Just by listening, humans can estimate object material, size, and impact force. The physics of how these parameters interact to generate sound is well established, yet the inverse problem faced by listeners – inferring physical parameters from sound – remains poorly understood. One challenge is that two objects create a sound, leaving the inference of any single object's properties ill-posed. Here we show that judgments of physical variables exhibit the phenomenon of explaining away, as might be expected if listeners perform inference in a generative model of acoustical physics. We presented listeners with recordings of a ball hitting a board. Listeners could identify the heavier of two balls, even when the two balls were dropped on boards differing in material – an ill-posed inference. To test whether listeners were implicitly estimating and discounting the board material, we altered the decay rates of the boards to imply a harder or softer material. These alterations affected mass comparisons despite not being directly informative about mass, suggesting that listeners factorize the acoustic contributions of the two objects. The results indicate that humans have internalized a generative model of impact sounds and use it to perform intuitive physical inference.

Keywords: intuitive physics; perception; natural statistics

Generative Model of Impact Sounds

How physical objects create sound is a classic physics problem (Helmholtz, 2013; Rayleigh, 1896). It is well established that objects vibrate most strongly at certain resonant frequencies known as "modes" (Fletcher & Rossing,1991; Morse & Ingard, 1968), the frequencies, powers, and decay rates of which are determined by the object shape and material. The sound of a ball striking a board can be modelled as

$$s(t) = \sum_{m=1}^{M} a_m e^{-b_m t} \sin(2\pi f_m t)$$

where the summation is over resonant modes of the board and a_m, b_m and f_m are the amplitude, decay rate and frequency of the mth mode. This assumes the ball radiates negligible sound which, unless the ball is hollow, is usually true.

We present theoretical arguments and experimental data to demonstrate that mode frequencies and decay rates are not affected by the ball. Hence, they provide explicit information about the board properties. The mode amplitudes, however, are jointly determined by the ball and the board. Thus, inference of the ball properties from an impact sound is fundamentally ill-posed. One possible solution is to use the explicit information about board material (i.e. b_m and f_m) to estimate material. The board contributions to mode power can then, in principle, be estimated and explained away (Fig 1). Such an inference would yield a better estimate of ball properties than an inference algorithm that considered ball and board properties independently.

Perceptual Experiments

We sought to test whether humans can infer ball properties when the board is unknown (Exp 1). To test whether humans employ a generative model to interpret sound, we tested whether their judgments of the ball are explicitly affected by changing mode decay rate (Exp 2), which is affected by board material but not by the ball properties (Fig 1).

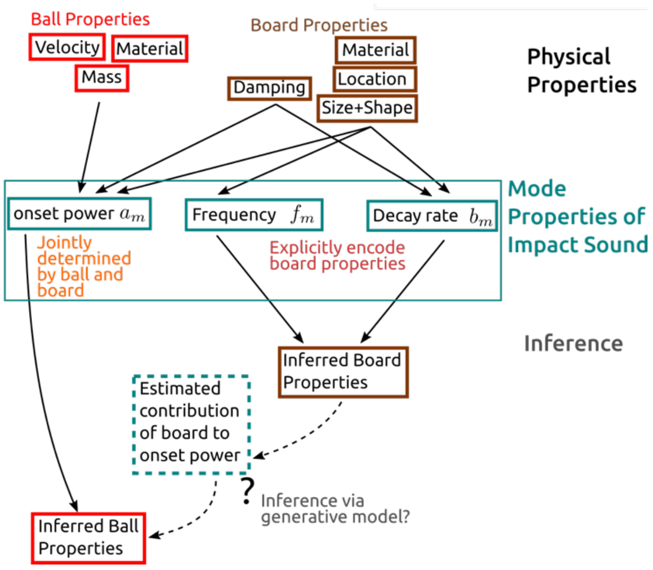

Figure 1: Generative model of impact sounds as a function of ball and board properties. We sought to test whether humans used the generative structure to explain away material contributions of the board to better infer the properties of the ball.

Experiment 1: Disambiguation of Board and Ball

We presented listeners with pairs of recorded impacts from different boards – one struck with a small ball and one with a large ball. Listeners were asked to judge which of the two balls was heavier, the first or the second (Fig 2a). A second experiment used the same ball dropped from different heights. In this case listeners were asked which impact was more forceful (i.e. which ball was moving faster at the time of impact). The results (Fig 2b) show that listeners can successfully infer ball properties, and that they do so better than simple classifiers which use simple acoustic features to perform the task (e.g. by choosing the sound with higher amplitude, lower spectral centroid, softer attack, or longer decay). The acoustic classifiers performed poorly because of the variation in board materials – the bigger ball did not always make the louder, or lower frequency sound. The results suggest that humans have some ability to disambiguate acoustic contributions of the ball from those of the board.

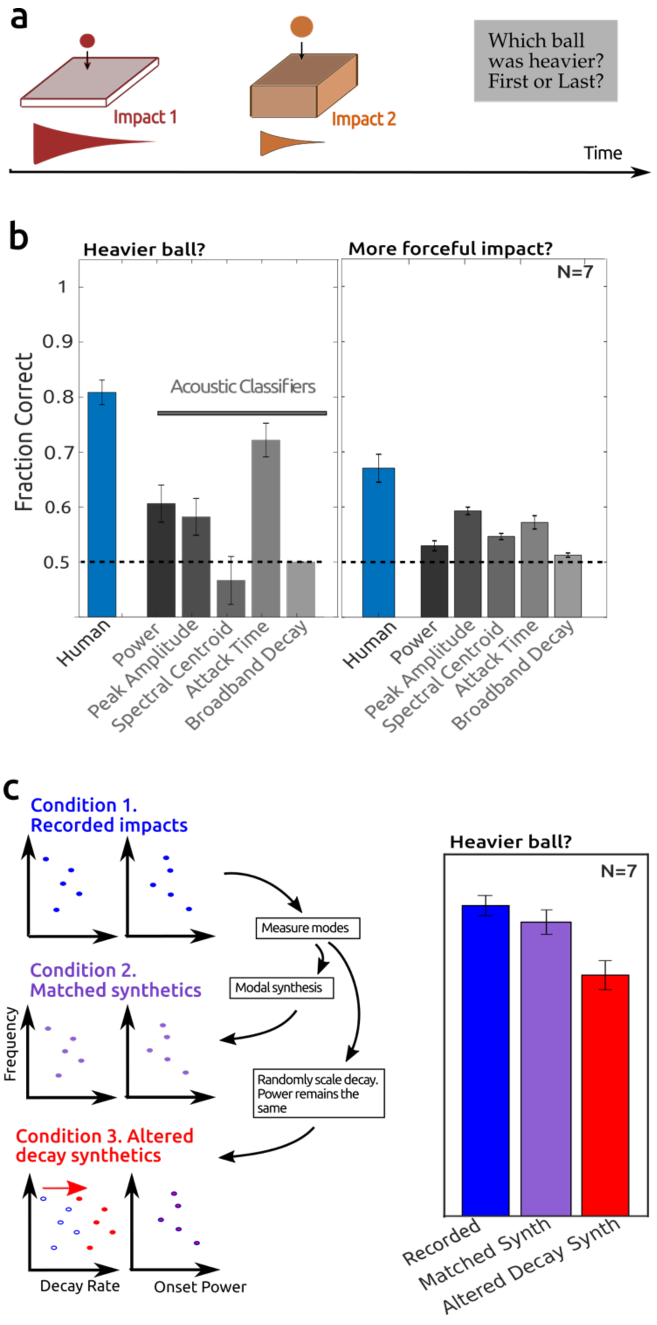

Figure 2: (a) The task was to select the heavier of two balls striking two different unknown boards (presented sequentially). (b) Results of human listeners (blue) and a battery of acoustic classifiers (grey) when estimating (left) the larger ball mass and (right) the more forceful impact. (c) Results of humans on the ball mass estimation task with recorded impacts, and synthetic impacts with parameters estimated from the recordings, or with altered decay rates (implying a different material).

Experiment 2: Explaining Away the Board

To test if humans explain away the board to better infer properties of the ball, we ran the same task as in Experiment 1, but used both recorded and synthetic sounds, created by summing decaying sinusoids (Klatzky et al. 2000; Lutfi et al. 2005; Aramaki & Kronland-Martinet, 2006; Giordano & Avanzini, 2014). Two classes of synthetic sounds were used. The first (matched synthetic) used modes matched explicitly to those of the recorded impacts. These were almost indistinguishable from the real-world recordings. The second class (altered decay synthetic) contained modes matched to real-world recordings but with the decay rates randomly re-scaled up (implying a softer material) or down (implying a harder material). The frequencies and onset powers remained fixed across both classes of synthetic sounds.

Humans made similar judgments of ball mass with the matched synthetics as they did with the recorded impacts (Fig 2c). However, when presented with synthetic sounds with altered decay times, listeners performed worse at the ball mass estimation task. This demonstrates that listeners take mode decay rate (a cue to board material) into account when judging ball properties.

Conclusion

We have shown that human listeners can discriminate small and large balls, (and light and forceful impacts), even when they strike boards of different materials. Moreover, humans alter their judgments of ball properties when board material cues are altered. The results suggest that listeners infer the material of objects and explain away effects of one object (the board) to better infer properties of the other (the ball). Humans have evidently internalized a generative model of impact sounds and use this knowledge to perform intuitive physical inference to infer the causes of sound in the world.

Acknowledgments

The authors are grateful to Nathan Munet, Sarah Royka and Jasmine Zou for help with audio recordings. This work was funded by a McDonnell Scholar Award.

References

Aramaki, M., & Kronland-Martinet, R. (2006). Analysis-synthesis of impact sounds by real-time dynamic filtering. IEEE Transactions on Audio, Speech, and Language Processing, 14(2), 695-705.

Fletcher, N. H., & Rossing, T. D. (2012). *The physics of musical instruments*. Springer Science & Business Media.

Giordano, B. L., & Avanzini, F. (2014). Perception and synthesis of sound-generating materials. In Multisensory Softness. Springer, London.

Helmholtz, H. (2013). *On the sensations of tone*. Courier Corporation.

Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of material from contact sounds. Presence: Teleoperators & Virtual Environments, 9(4), 399-410.

Lutfi, R. A., Oh, E., Storm, E., & Alexander, J. M. (2005). Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians, and nonmusicians. The Journal of the Acoustical Society of America, 118(1), 393-404.

Morse, P. M., & Ingard, K. U. (1968). *Theoretical acoustics*. Princeton university press.

Rayleigh, J. W. S. B. (1896). *The theory of sound* (Vols. 1 & 2). Macmillan.