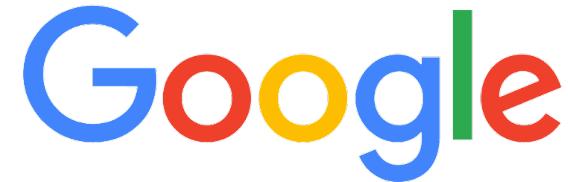
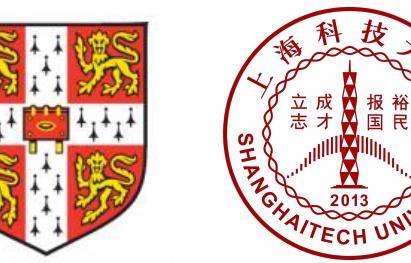


Shape and Material from Sound

Zhoutong Zhang¹, Qiuja Li², Zhengjia Huang³, Jiajun Wu¹, Joshua B. Tenenbaum¹, William T. Freeman^{1,4}

¹Massachusetts Institute of Technology, ²University of Cambridge, ³ShanghaiTech University, ⁴Google Research



Motivation

Can you tell the shape & material from sound?



Harder!



Problem:

- Computational inference from sound?

Human may use a mental physics engine;

Inferring with an audio engine: reconstruct input

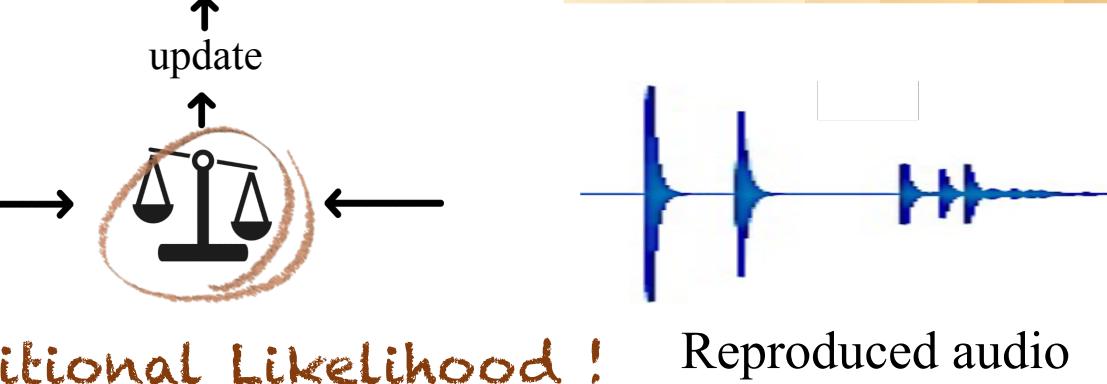
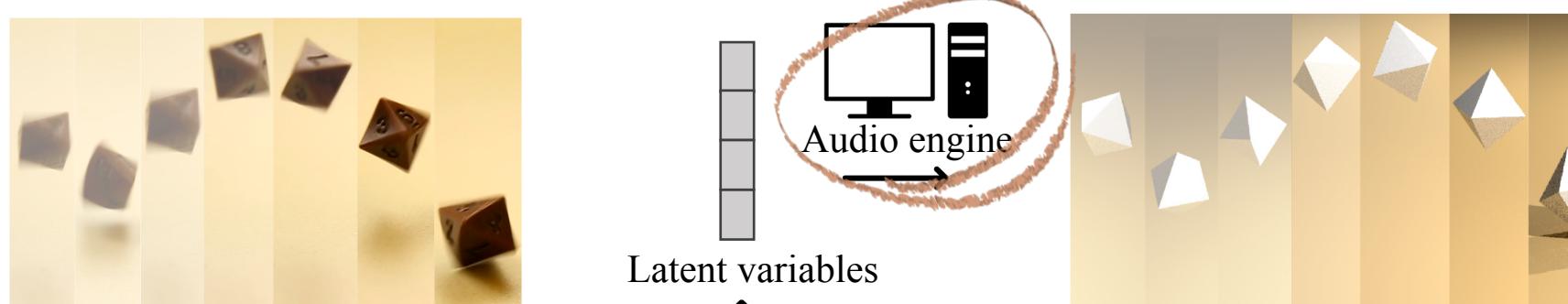
Challenge:

- Fast audio engine?

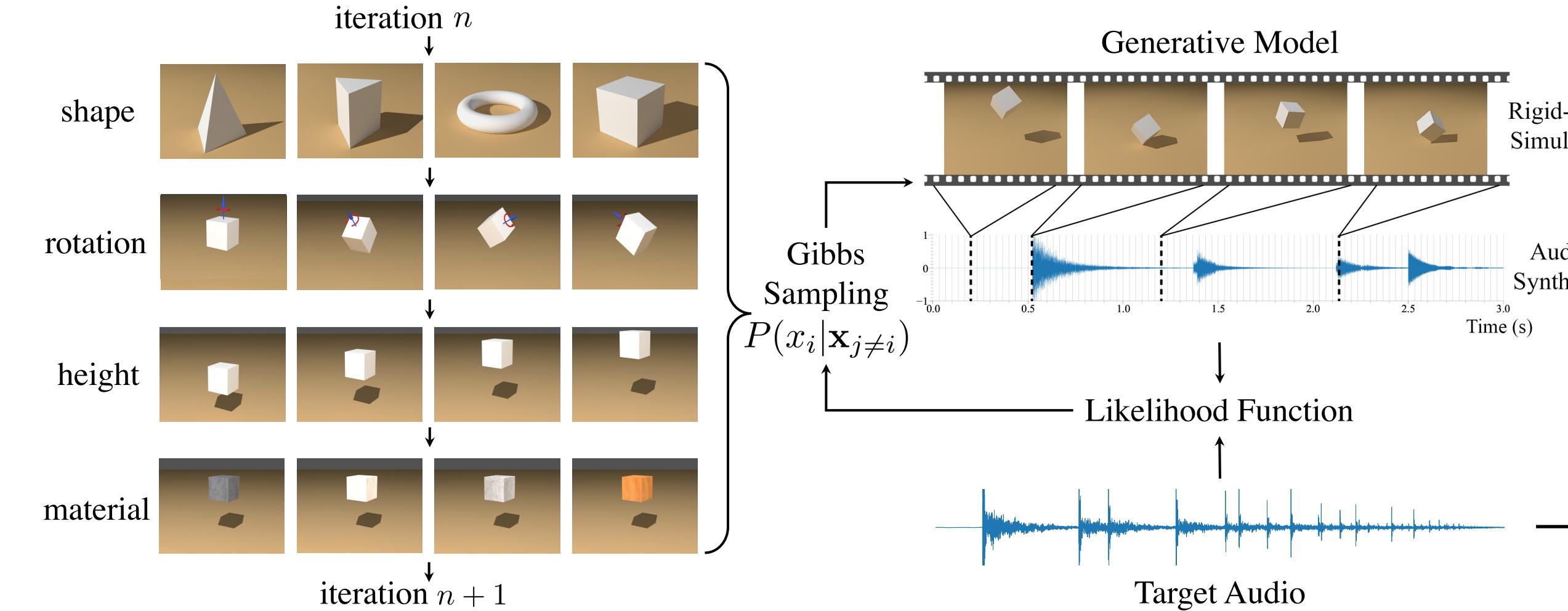
- Measuring similarity between audio?

Solution:

Near real-time!

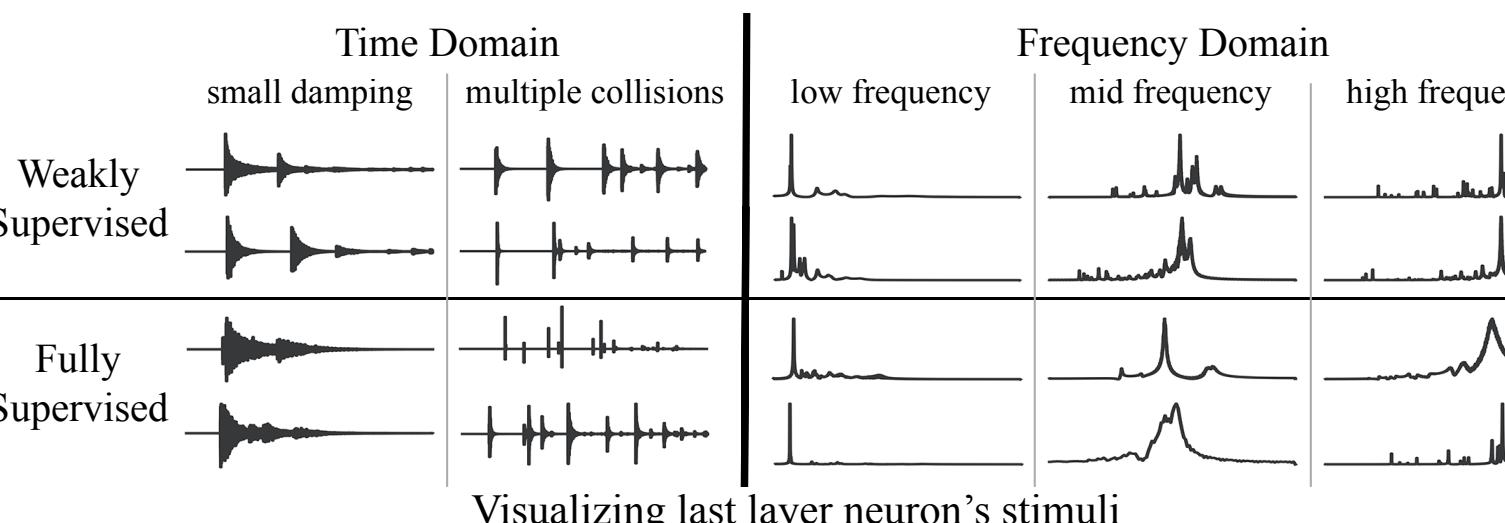


Approach



Results: Comparing models

Inference Model	Latent Variables				
	shape	mod.	height	α	β
Unsupervised	init	8%	10%	0.179	0.144
	infer	54%	56%	0.003	0.069
Self-supervised	init	14%	16%	0.060	0.092
	infer	52%	62%	0.005	0.061
Weakly supervised	init	18%	12%	0.018	0.077
	infer	62%	66%	0.005	0.055
Fully supervised	infer	98%	100%	0.001	0.001
				0.051	

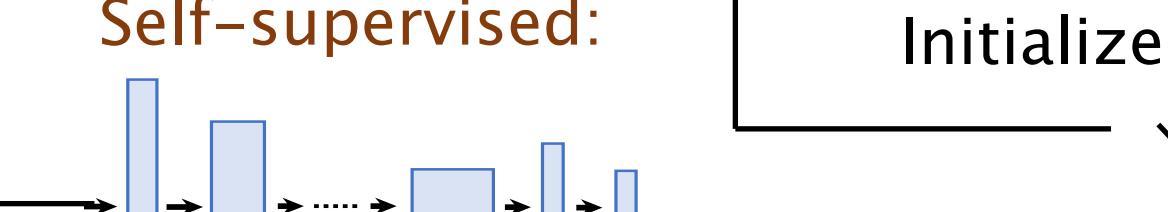


Models

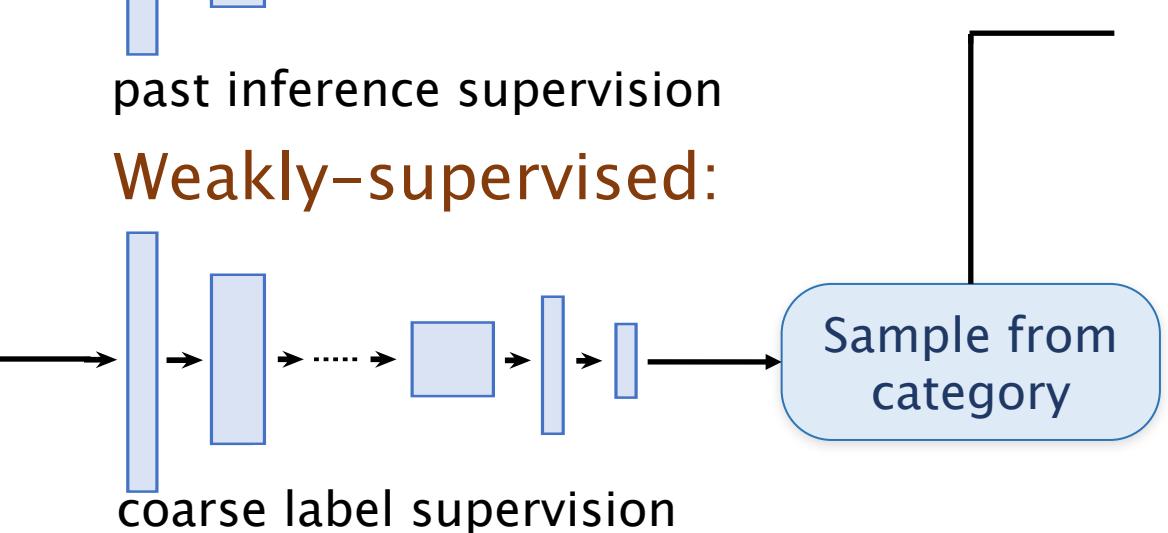
Unsupervised:

Random Sampling

Self-supervised:



Weakly-supervised:



Latent variables

Results: Comparing with Human

