CSE 421

ARTIFICIAL INTELLIGENCE

REPORT ON

Shape and Material from Sound

Submitted By:
Prerana DAS
ID:161-115-119
Session:2016-2019
Batch: 38th
Section: C

Instructor:
Arif AHMED
Senior LECTURER
Dept. of CSE

September 6, 2019

মেট্রোপলিটন ইউনিভার্সিটি
Metropolitan UNIVERSITY

Contents

List of Figures

1 Introduction

1.1 Main Idea

When One object fall onto the ground it's makes a sound. By hearing this sound any human being can predict this object shape material and falling height. Because one human being has unknown capacity of knowledge.So at a short period of time they can make a big amount of information.
In this paper they build such a machine to predict all of this properties.

- Estimate physical properties of the objects
- Compare this predictions with human response
- The developed model makes to some real data
- The model achieves near-human performance

1.2 Contribution

Build a machine which can infer raw shapes, materials and falling height of an object from sound.
First of all we collect a set of human knowledge for an falling object by the help of physics based simulation engine. Also use analysis by synthesis approach to infer properties. Then map their prediction with previous information.

2 Major Experimental Results

Construct an audio dataset that includes 14 primitives .Each with 10 different moduli.

Variable	Range	C/T
Primtive Shape(s)	14 classes	D
Height (z)	[1, 2]	C
Rotation axis (i, j, k)	S2	C
Rayleigh damping (α)	$10[-8, -5]$	C
Specific modulus (E/ρ)	$[1, 30] \times 106$	D
Restitution (e)	[0.6, 0.9]	C
Rotation angle (w)	$[-\pi, \pi)$	C
Rayleigh damping (β)	2[0,5]	C

Initial and final classification accuracies and parameter MSE errors of three different inference models after 80 iteration of MCMC are showing below.

Gibbs sampling

1. Begin with some initial value $X^{(i)}$

2.We want the next sample. Call this next sample $X^{(i+1)} = (X^{(i+1)}, X^{(i+2)}, ...X^{(i+n)})$

3.Update it according to the distribution specified by $X_j^{(i+1)}|(X_1^{(i+1)}, ...X_j - 1^{(i+1)}, X_j + 1^{(i)}....X_n^{(i)})$

3 Strength And Weaknesses

3.0.1 Strength

No need to visual perception.Three binary judgemental about the shape by listening to our synthesized audio clip. Humans are relatively good at recognizing shape attributes from sound. From a audio clips we get some physical properties those are density young's modulus and damping coefficients. This model also can measure falling height whether an object dropped from to the ground. Find out the sence setup .

3.0.2 Weaknesses

For inferring materials from sound let one physical objects property has four possible materials those are steel, ceramic, polystyrene and wood.There should be choose one out of four parameter.However sometimes this is to much difficult when sampled ones have similar damping and specific modulus.
This machine confused steel with ceramic and ceramic with polystyrene. This is the greatest weakness in this inference model.

4 Follow-up Works

For sound wave simulator. This simulation lets you see the movement of a sound wave. Adjust the frequency and can see and hear how the wave change.

4.1 Unsupervised

Recover the latent variables x to make the reproduced sound. Use the spectrogram as a feature and measure the l_2 distance between the spectrograms of two sounds, because Young's modulus will only affect the frequency at each collision

4.2 Self-supervised Learning

- Train a deep neural network
- Limited number of iterations
- Runs analysis-by-synthesis algorithm to refine the inference

4.3 Weakly-supervised Learning

- It is more realistic to assume
- Limited number of data that easy to get
- Refer to learning with noisy labels

4.4 Fully-supervised Learning

- To visualize the abstraction and characteristic features
- Inputs maximally activated

$SoundSynthesis \rightarrow SoundPropagation \rightarrow SoundRendering$

List of Figure

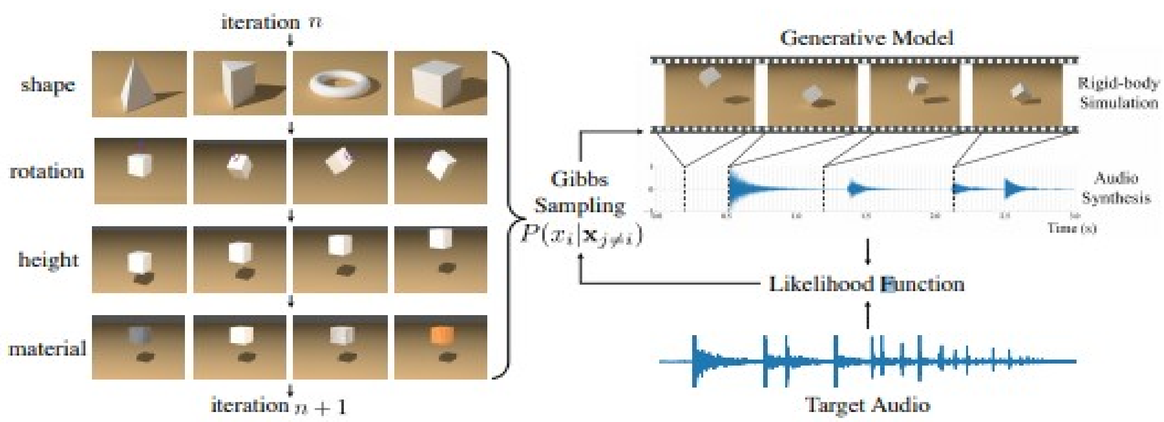

Figure 1: Our inference pipeline. We use Gibbs sampling over the latent variables. The conditional probability is approximated using the likelihood between reconstructed sound and the input sound

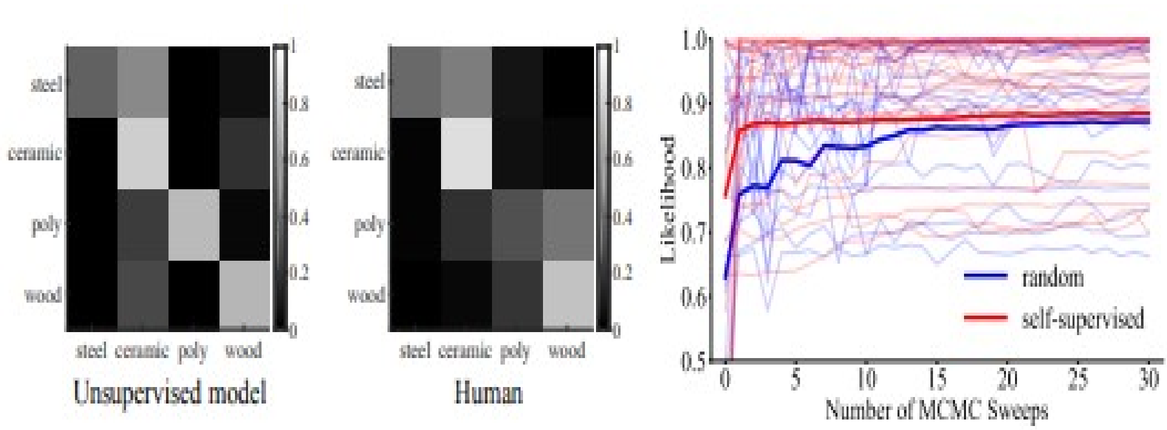

Figure 2: Left and middle: confusion matrix of material classification performed by human and our unsupervised model. Right: mean likelihood curve over MCMC iterations

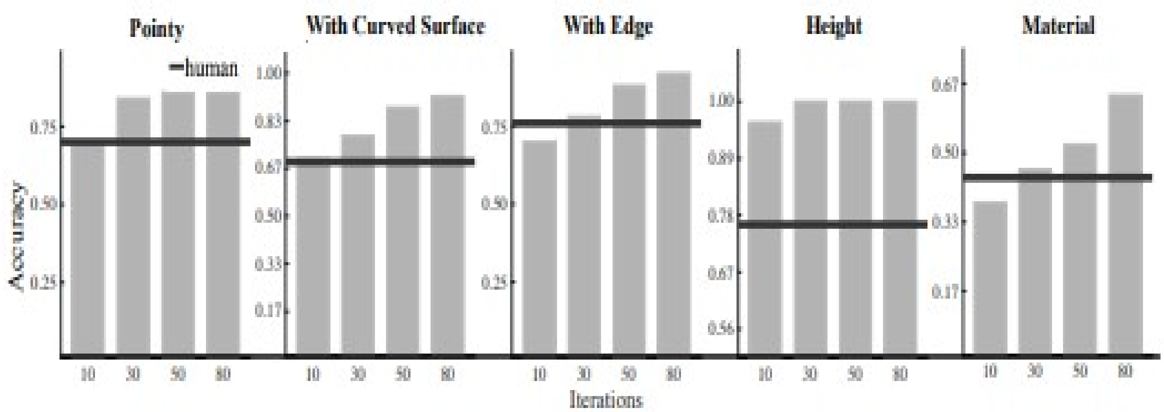

Figure 3: Human performance and unsupervised performance comparison. The horizontal line represents human performance for each task. Our algorithm closely matches human performance

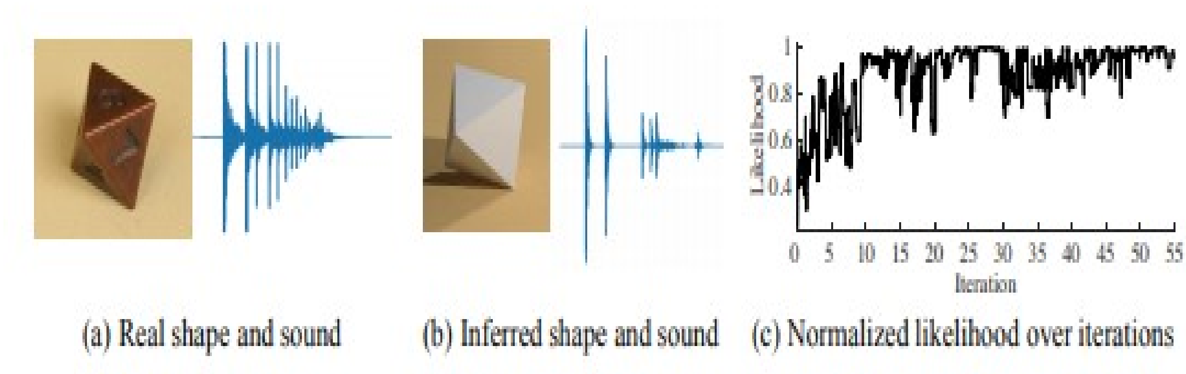

Figure 4: Results of inference on real world data. The test recording is made by dropping the metal dice in (a). Our inferred shape and reproduced sound is shown in (b). Likelihood over iteration is plotted in (c)

[1]

References

[1] Qiujia Li Zhoutong Zhang. *More math into latex.* 31st Conference on Neural Information Processing Systems, 2017.