Visual Vibrometry:
Using Cameras as Displacement Sensors

Miki Rubinstein

Google Research, Cambridge MA

(Work done while I was at MIT, Microsoft, Google)

Google Cambridge MA

Google Cambridge
(~1000 people)

MIT Stata Center (CSAIL)

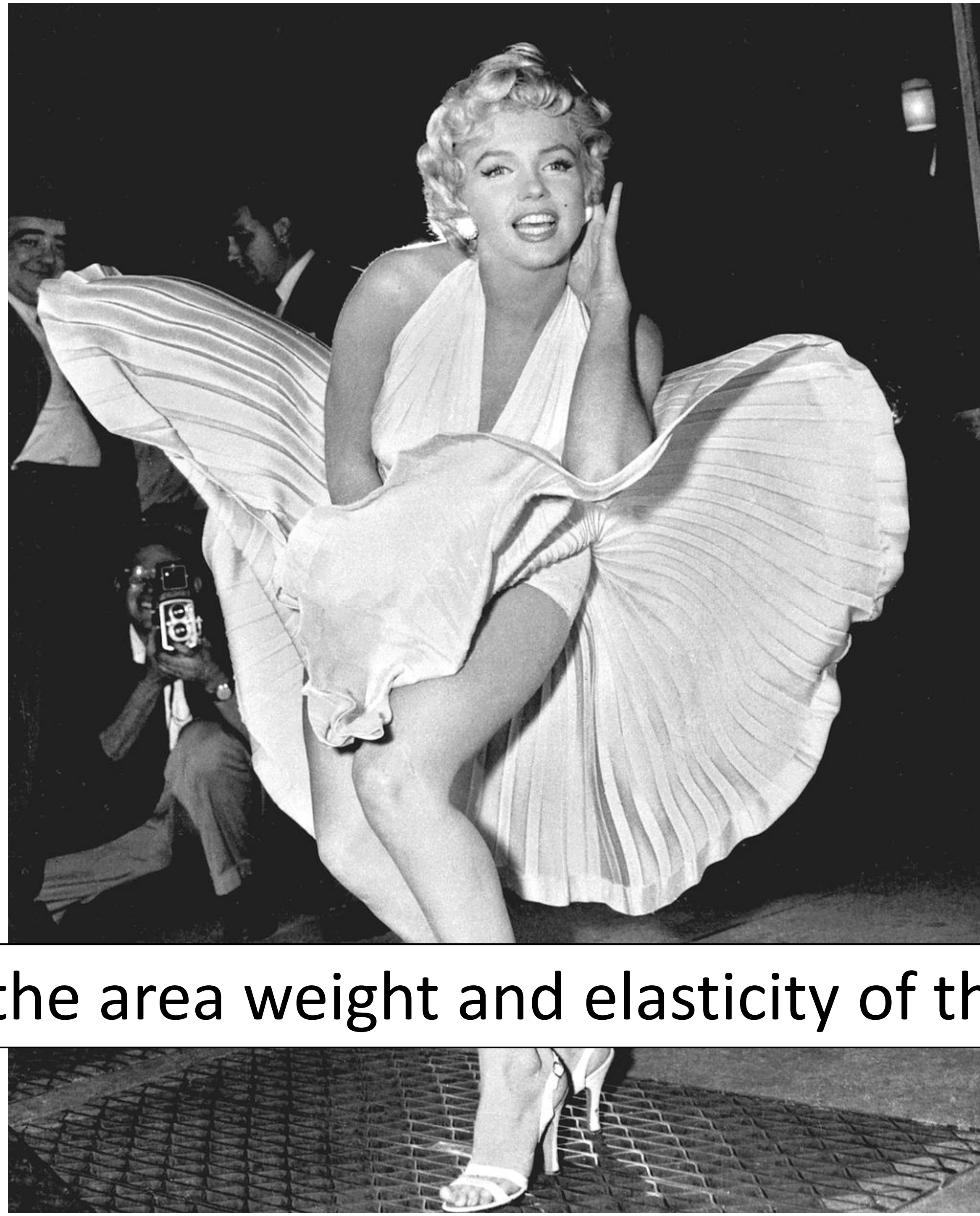
Google Cambridge Vision Group

- Started 2015, working at the intersection of computer vision and computer graphics.

- Group members:

Bill Freeman (MIT), Ce Liu (MIT), Miki Rubinstein (MIT), Dilip Krishnan (NYU), Forrester Cole (Princeton), Inbar Mosseri, Aaron Sarna

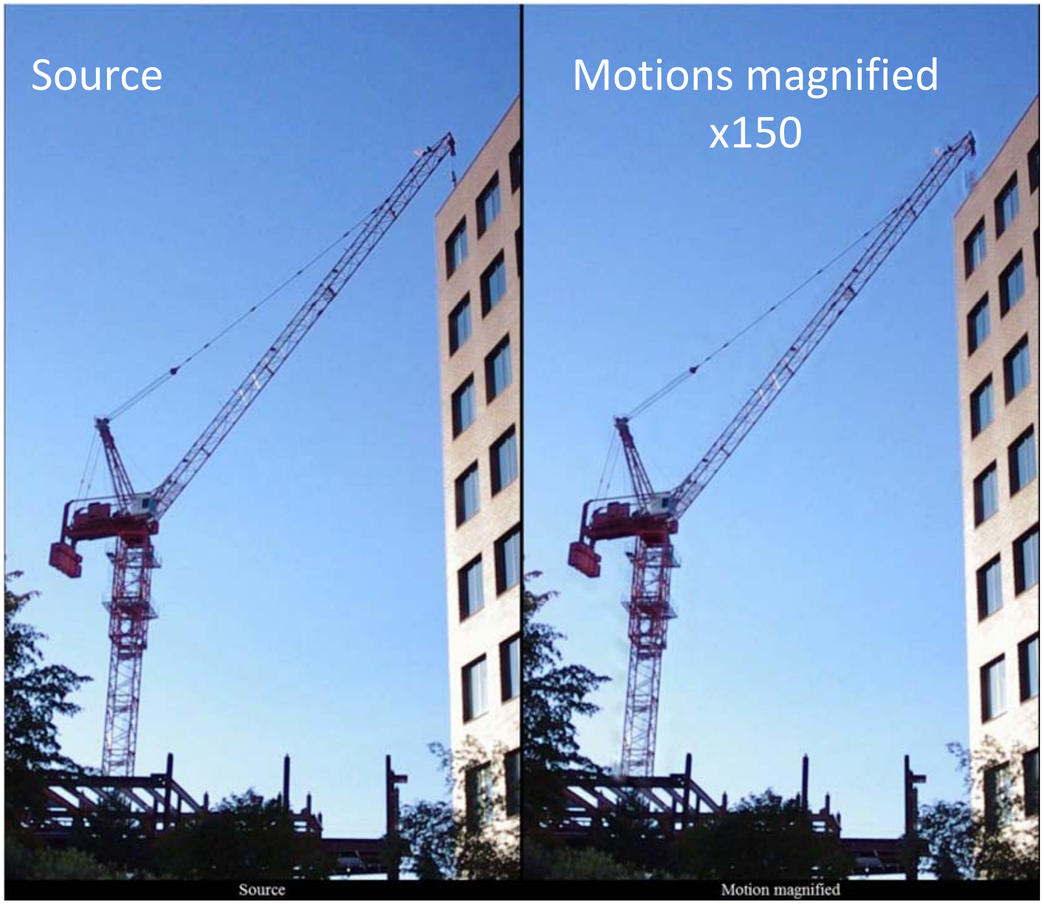
- Hiring student summer interns



What is the area weight and elasticity of this dress?

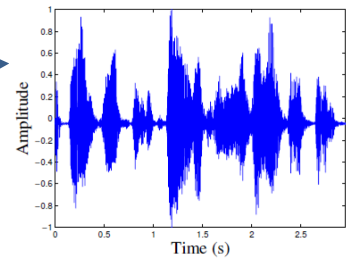
Tiny Motions in Videos

The "Motion Microscope"

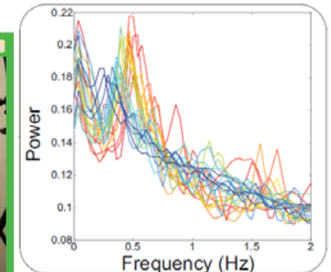


Not Just for Visualization

This talk

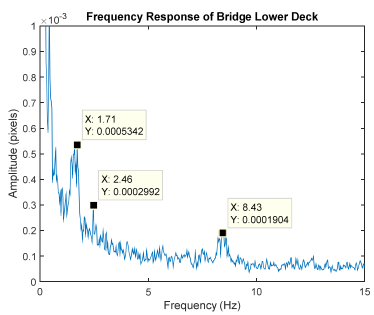


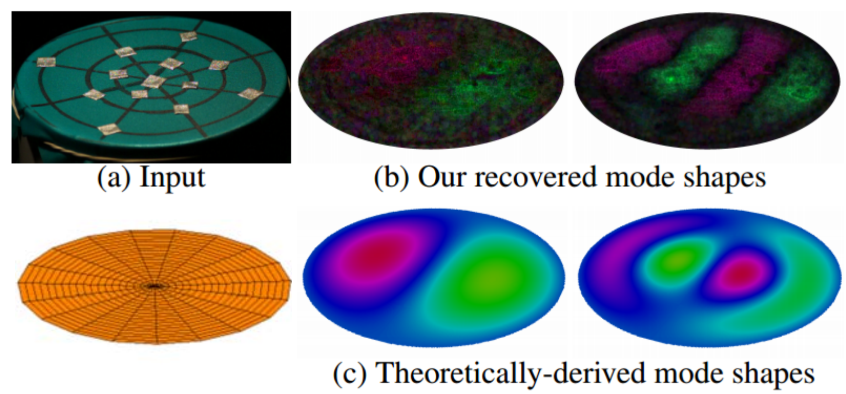
Recovering sound near objects



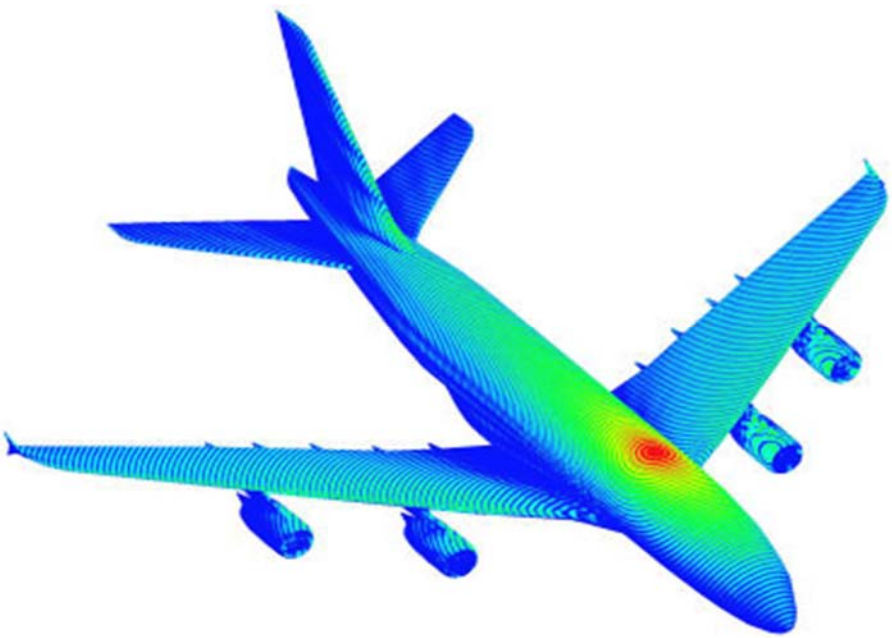
Estimating material properties
(weight, stiffness)

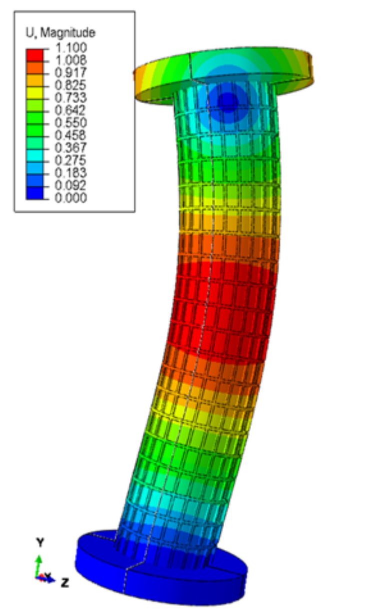
Extracting modal frequencies, mode shapes





Vibrometry, Non-Destructive Testing





Common Vibrometry Tools

Contact Sensors, accelerometers

Laser Vibrometers

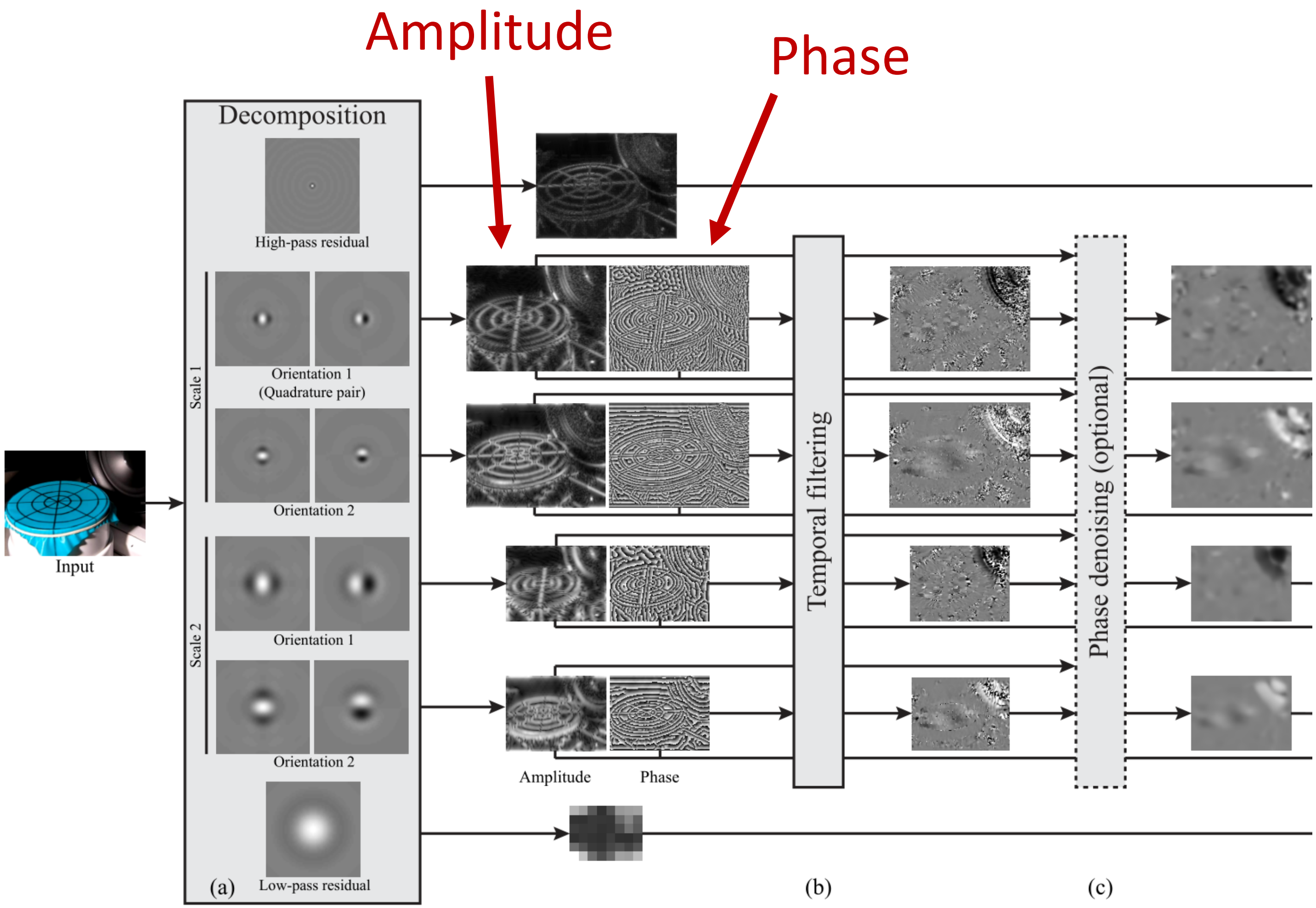
Vibrometry with Cameras

- **Small motions in videos: hard to see, easier to analyze**

- **Constraints**:
  - Textured surface,
  - Sufficient lighting,
  - Need high-speed camera for high-frequency motion

- **Advantages:**
  - Passive,
  - Non-contact,
  - Often more accessible / cheaper,
  - ***Spatial resolution!***

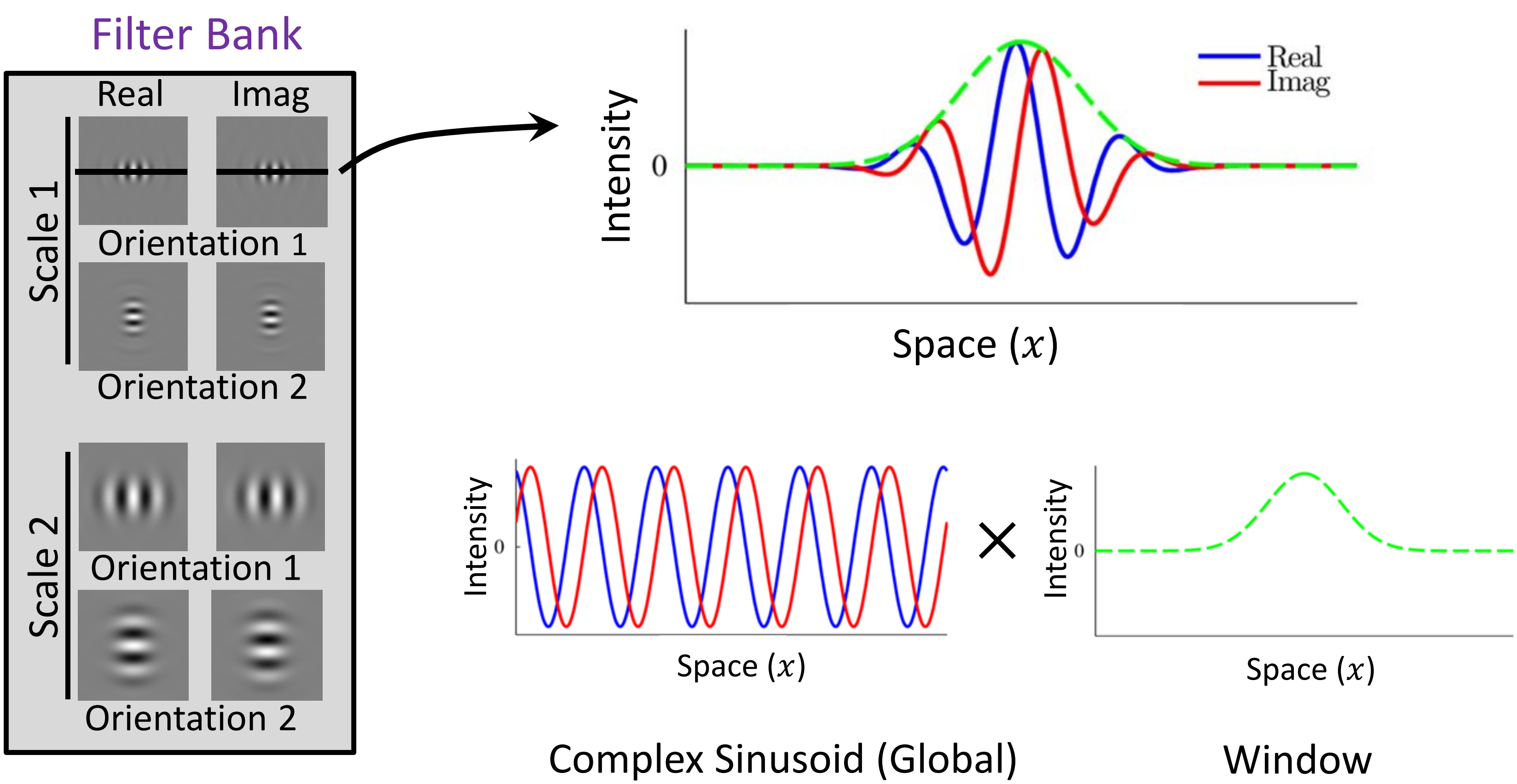
Phase-based Motion Processing (SIGGRAPH 2013)



Complex steerable pyramid
[Simoncelli and Freeman 1995]

Complex Steerable Pyramid [Simoncelli and Freeman 1995]

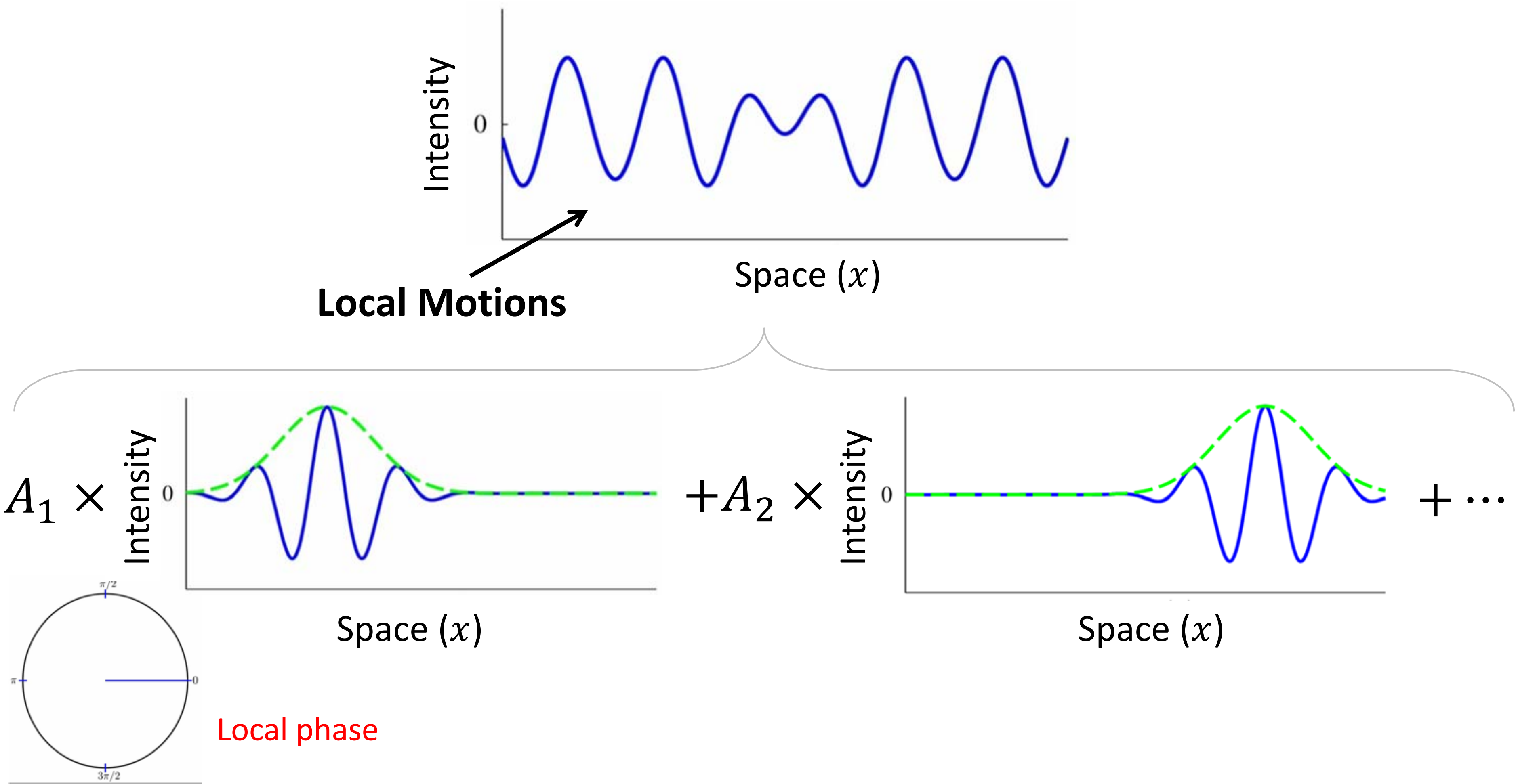
- Basis functions are wavelets with even (cosine) and odd (sine) components which give local amplitude and phase



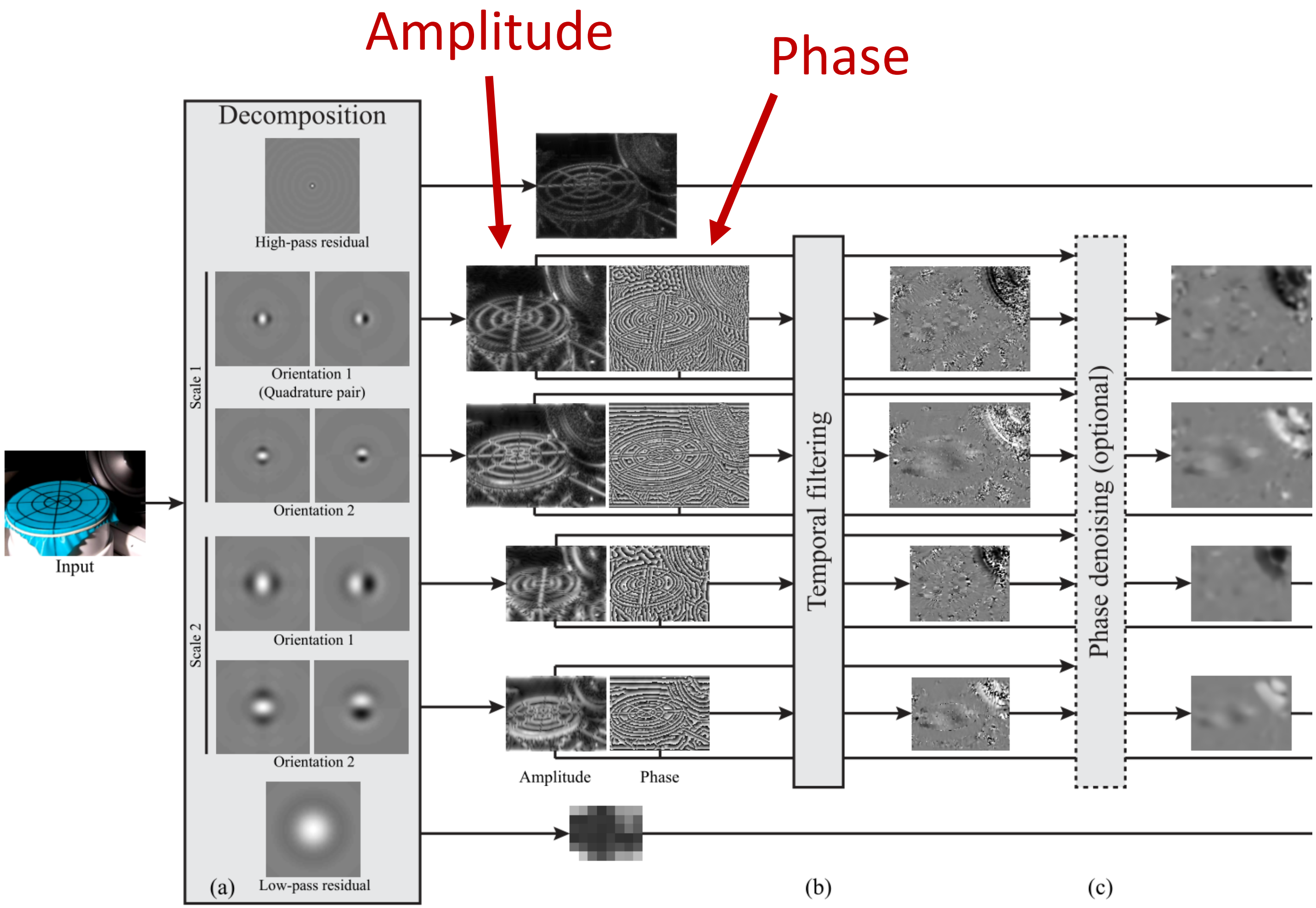
Local Phase

Local phase shift ↔ Local translation

- In a single subband, image is coefficients times translated copies of basis functions



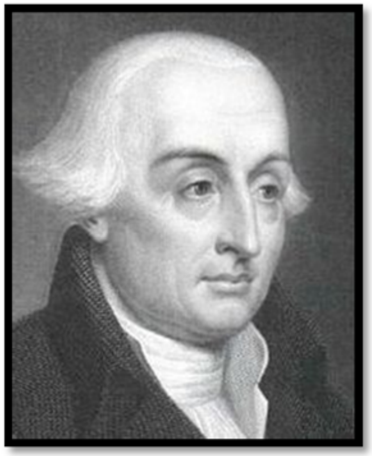
Phase-based Motion Processing (SIGGRAPH 2013)

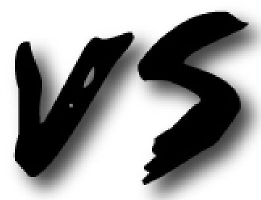


Spatial decomposition (representation) → Temporal filtering (per pixel)

Eulerian vs. Lagrangian processing (Fluid Dynamics)

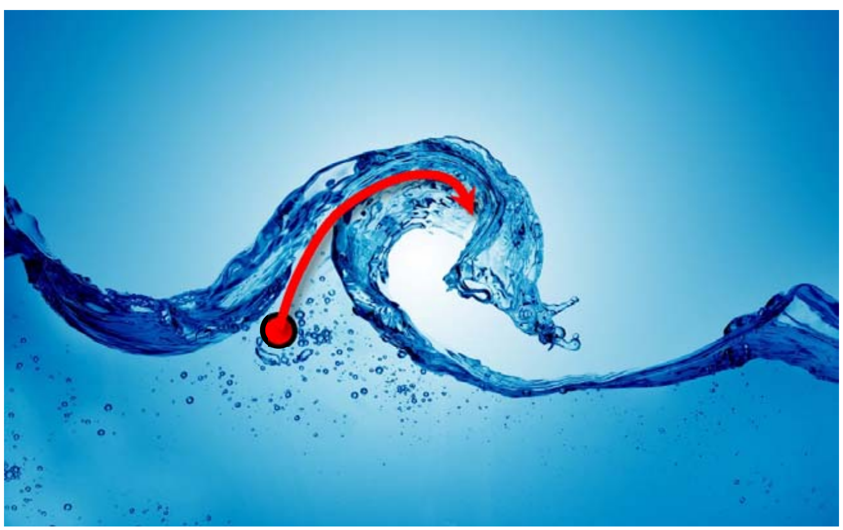
Joseph Louis Lagrange



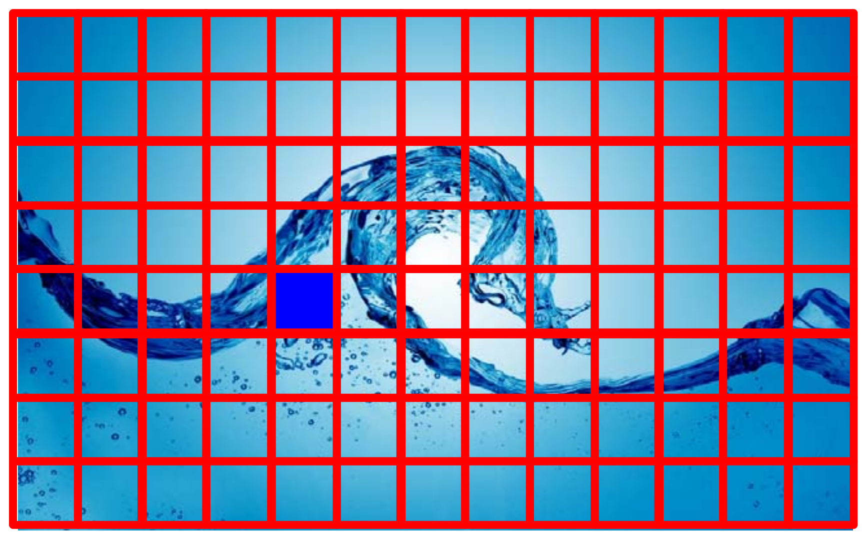


Leonhard Euler

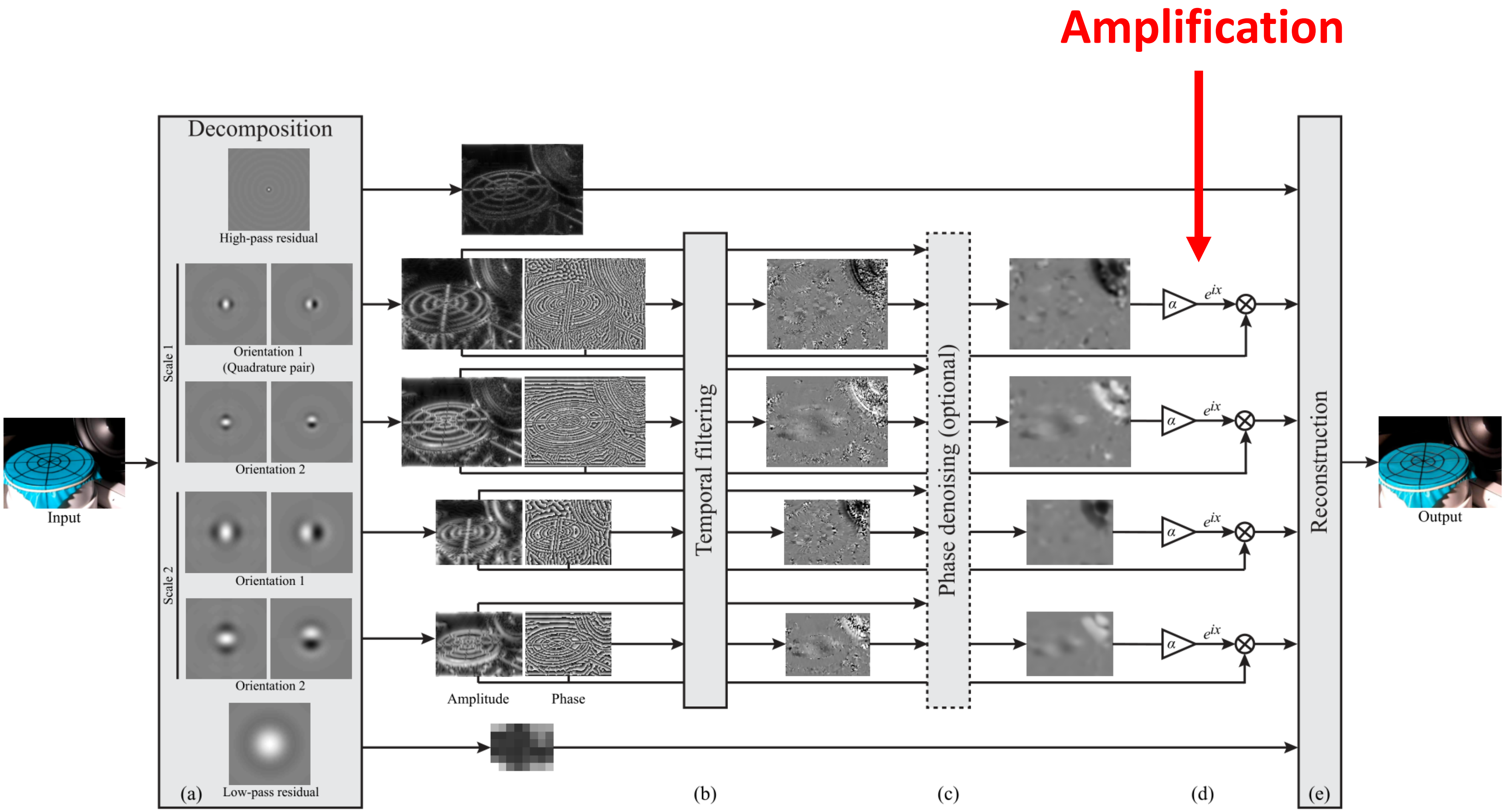
Track particles



Measure changes within fixed voxels in space



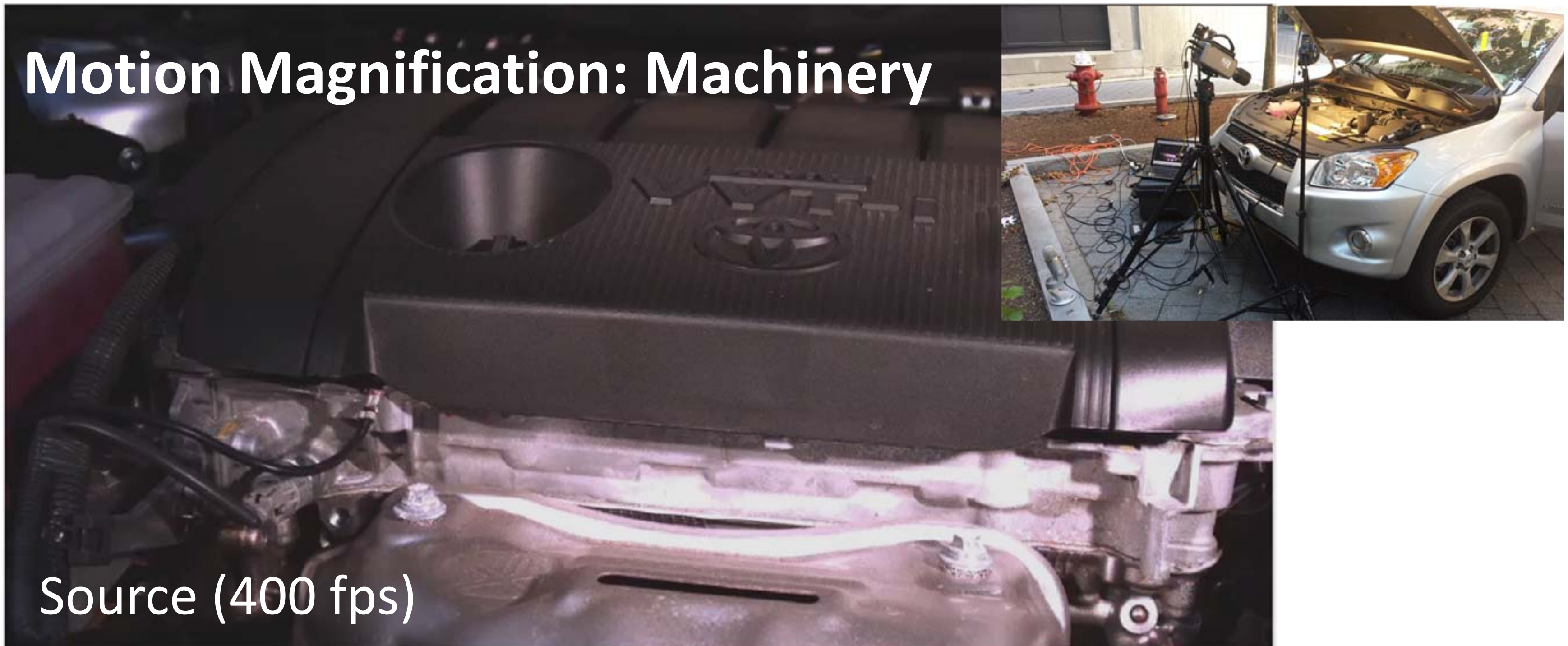
Motion Magnification



Spatial decomposition (representation) → Temporal filtering (per pixel) → Amplify variation (per pixel)

Motion Magnification: Physiological Signals

Motion Magnification: Machinery



Motion Magnification: Vibration Modes

Source (20000 FPS)

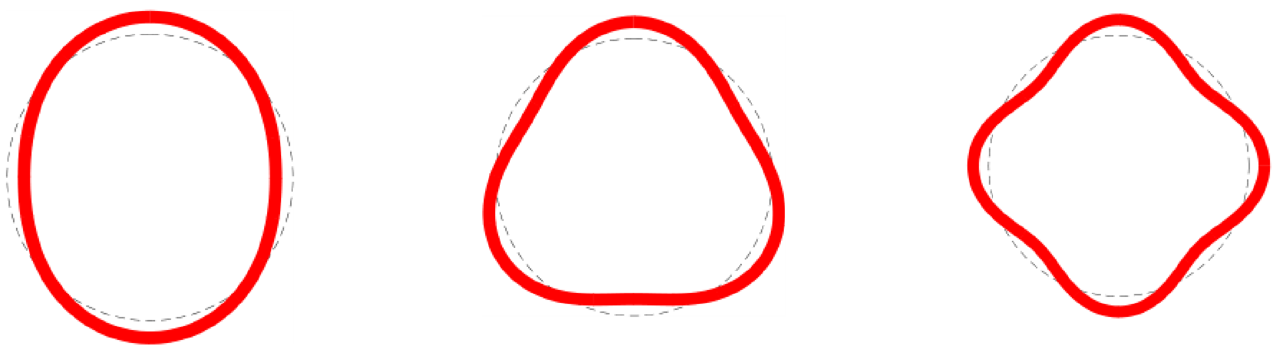
480Hz (x200)

1200Hz (x400)

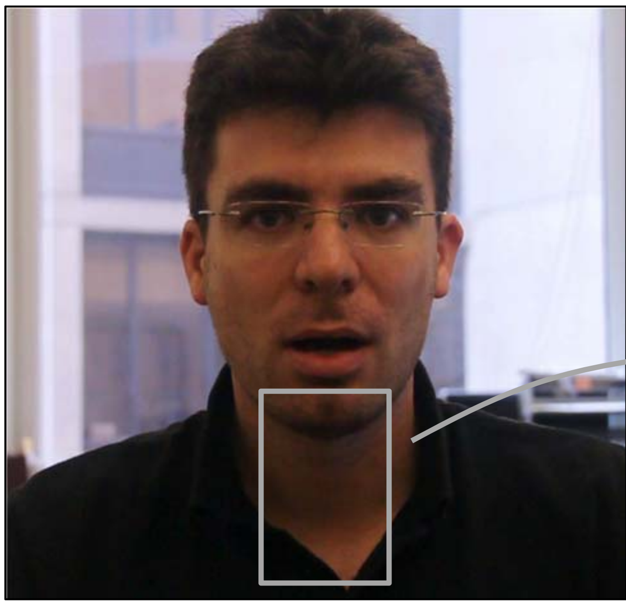
2400Hz (x1200)

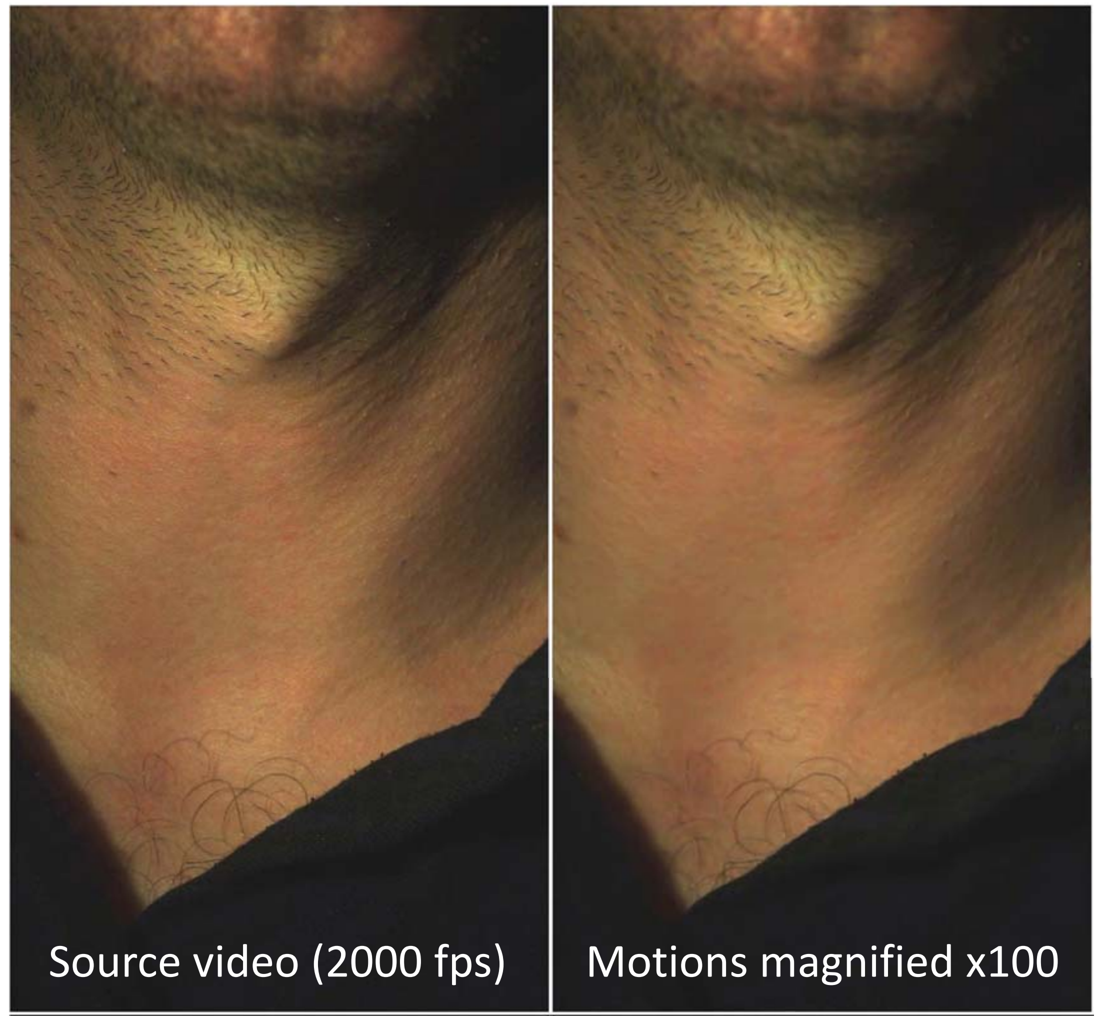
With Justin Chen, MIT Civil Engineering

Theoretically-Derived
Modal Shapes
[Wachel et al. 1990]



Sound and *Visual* Motion



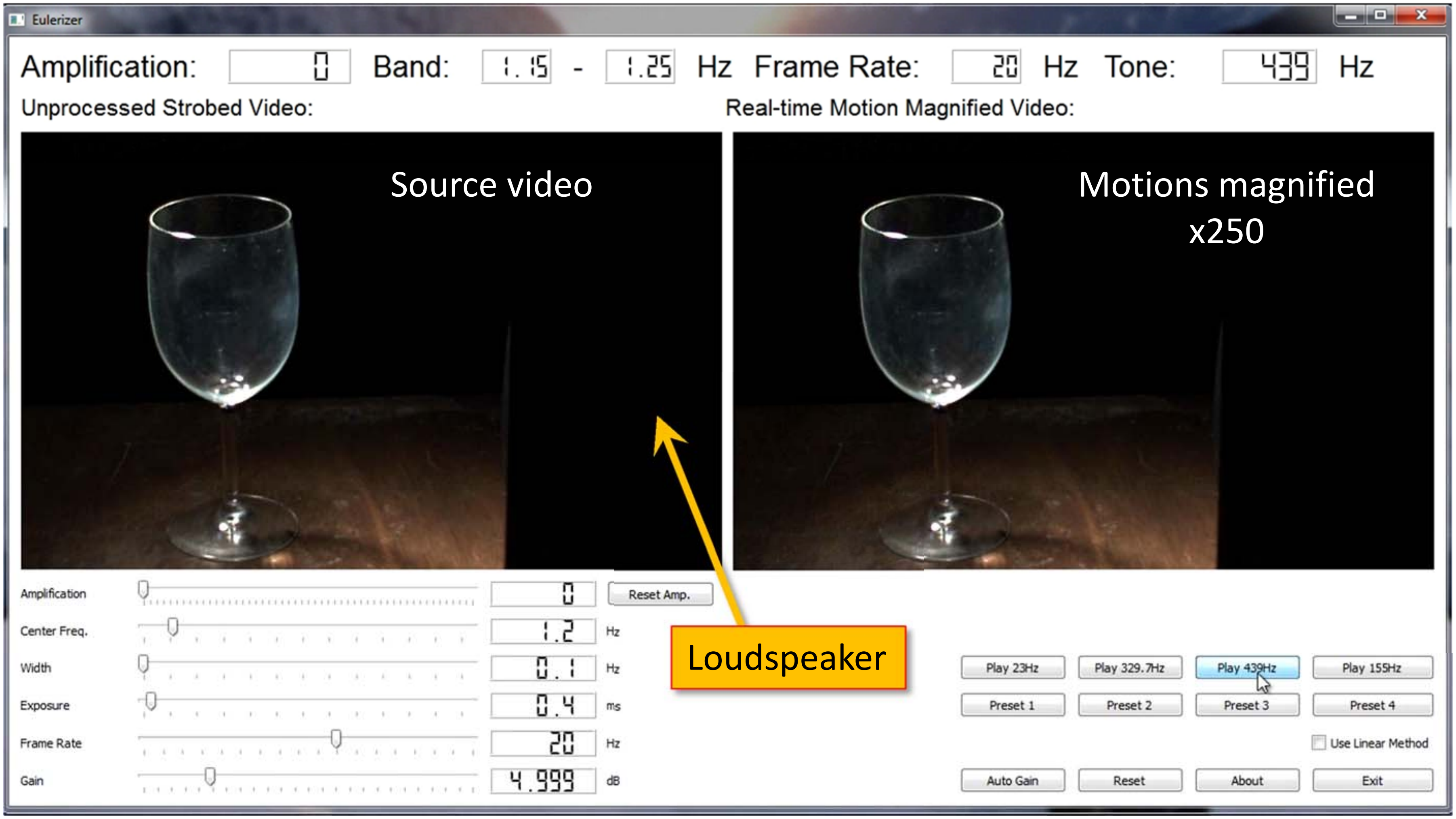


90-110Hz
(male voice pitch)

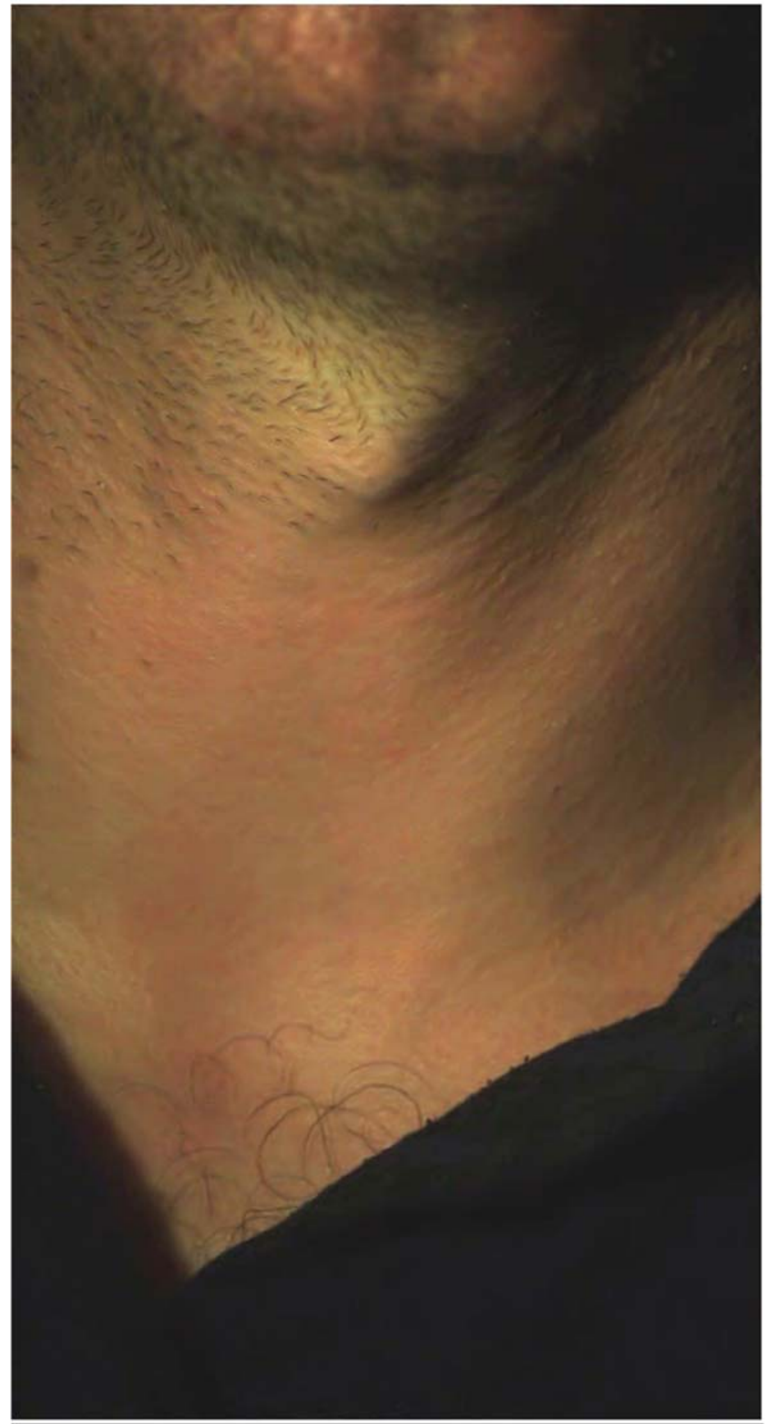
Sound and *Visual* Motion

Riesz Pyramids for Fast Phase-Based Video Magnification (ICCP'14)
With Neal Wadhwa, Fredo Durand, William T. Freeman

CVPR 2014 Best Demo Award



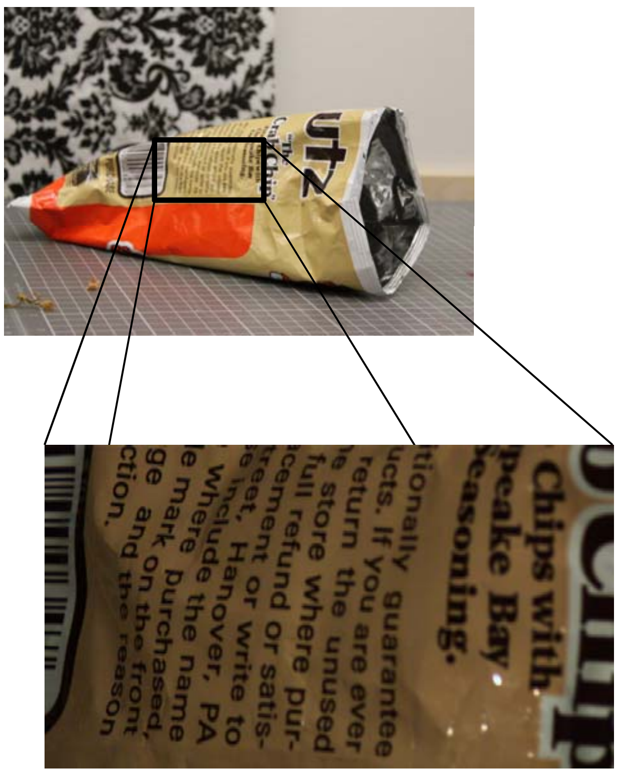
Sound-related Vibrations



Can We Recover Sound from Video?


"The Visual Microphone"
SIGGRAPH 2014
"The Crab Chip"
Potato Chips with Chesapeake Bay Crab Seasoning.
PURCHASE

Sound Recovered from Video

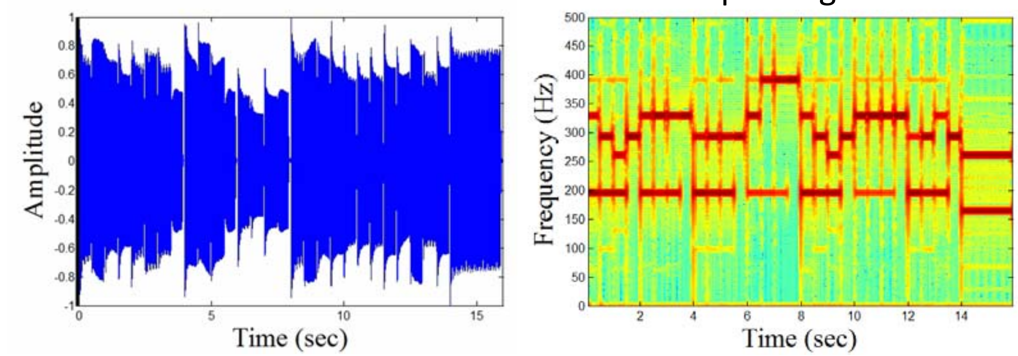


2200Hz (silent) video

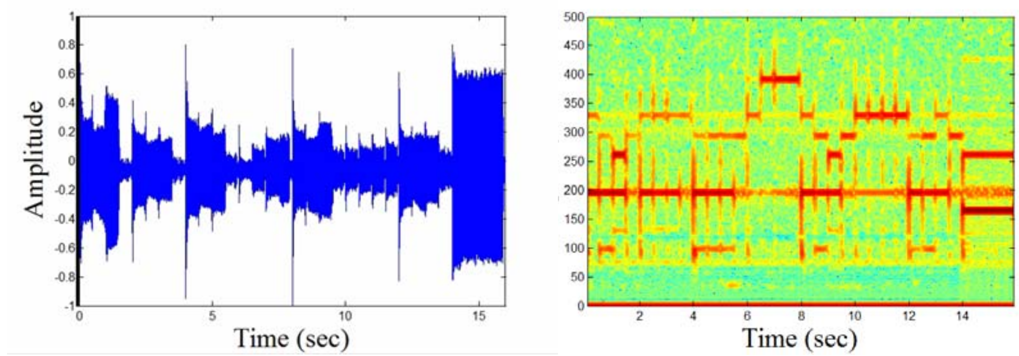
Source sound in the room

Waveform

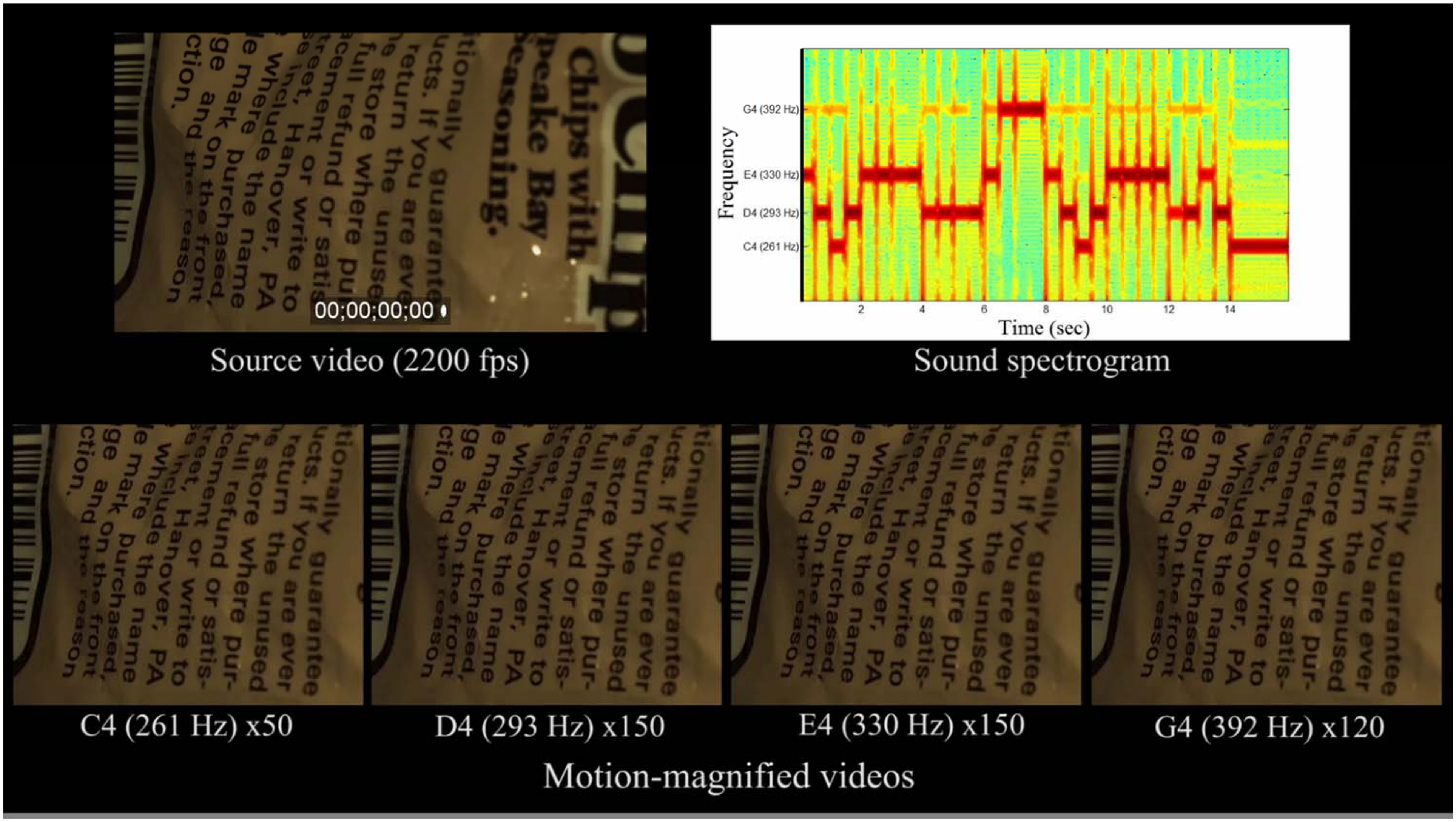
Spectrogram



Recovered sound



Motion-magnified Chip Bag

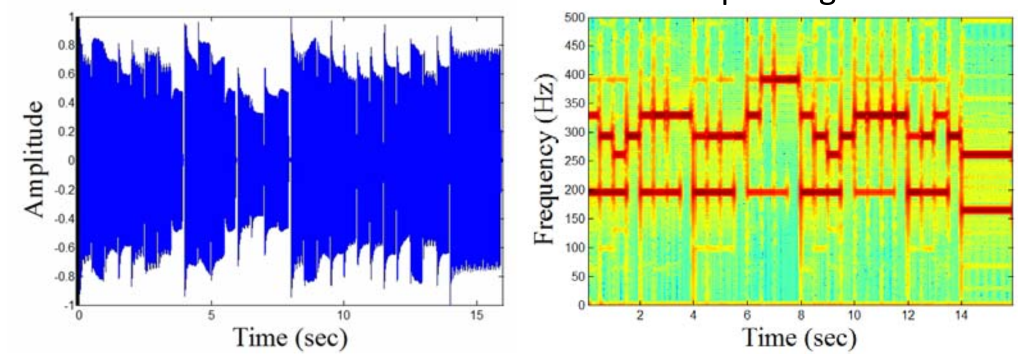


Sound Recovered from Video

Source sound in the room

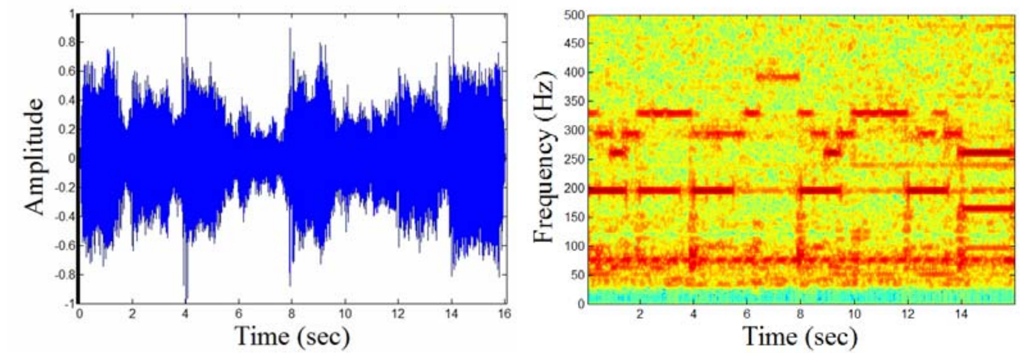
Waveform

Spectrogram

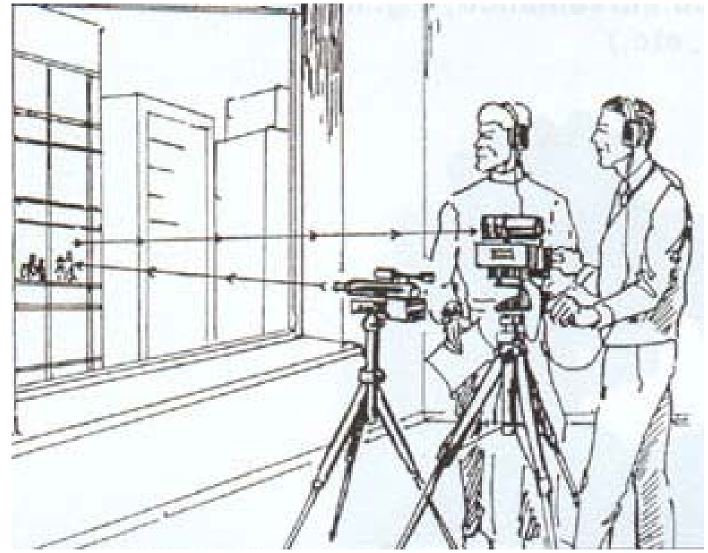


Recovered sound

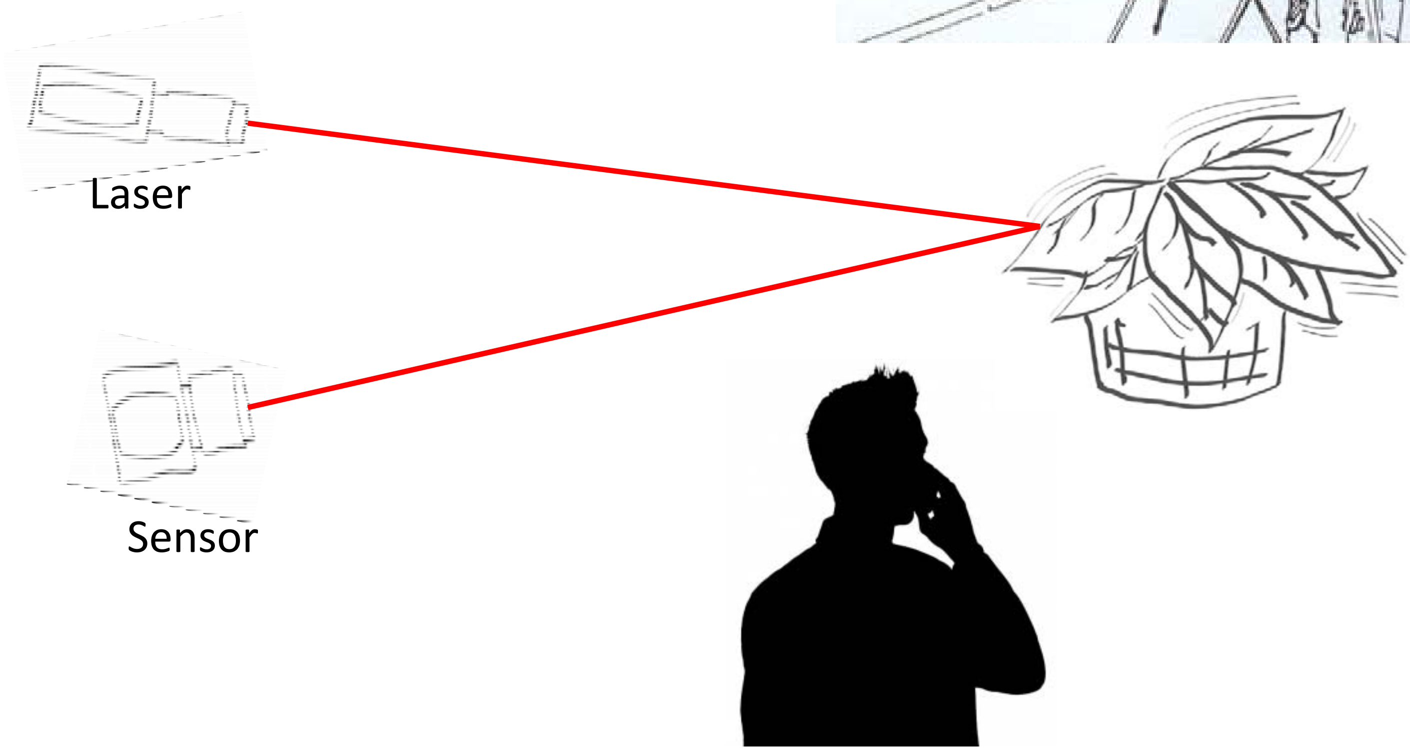
2200Hz video



Remote Sound Acquisition

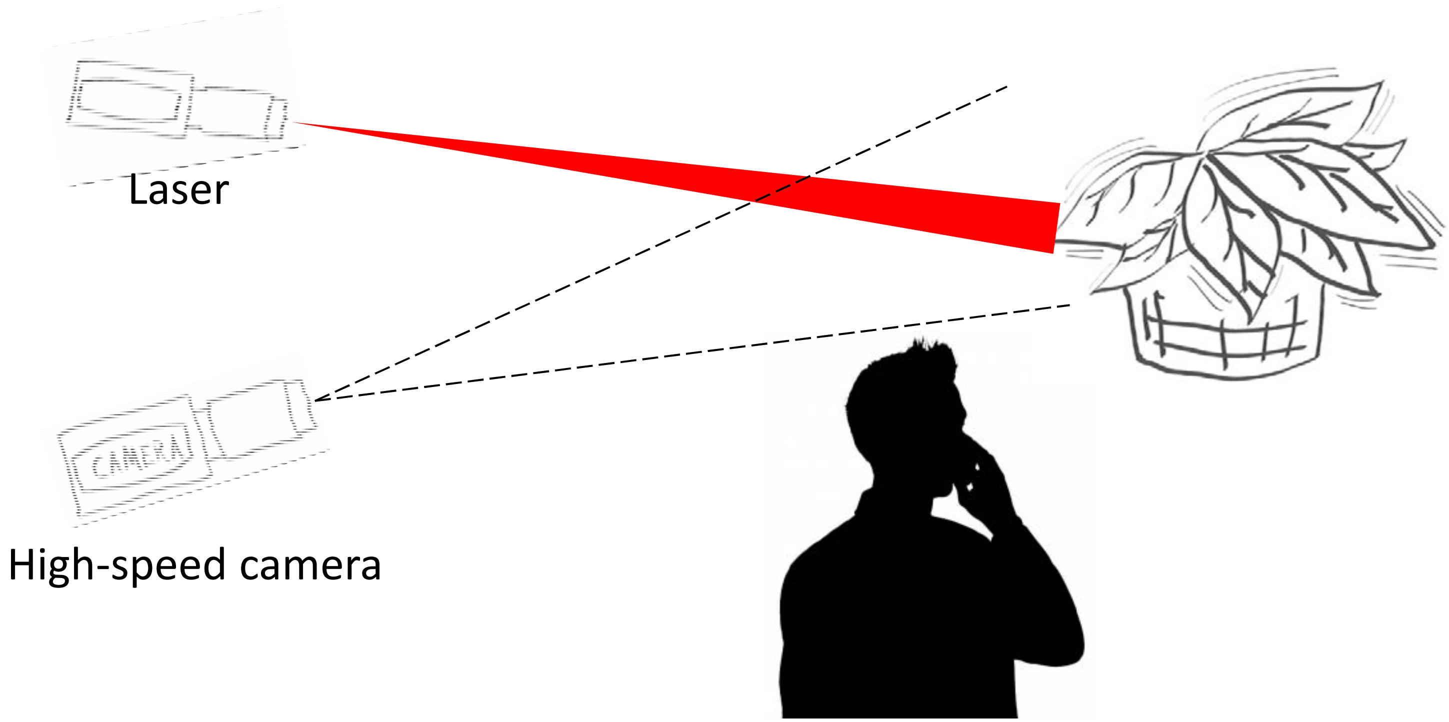


- *Active* techniques
  - Laser Microphone



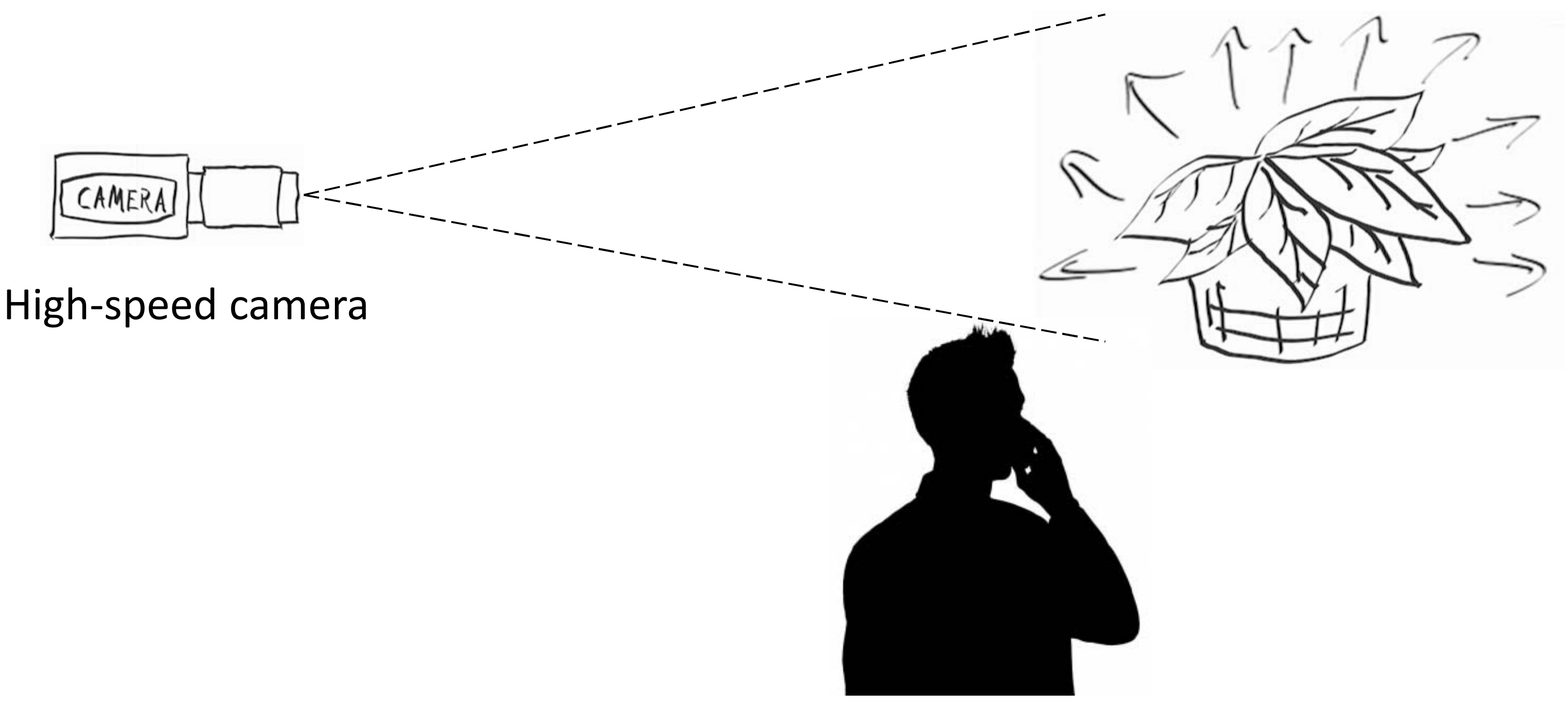
Remote Sound Acquisition

- *Active* techniques
  - Laser Microphone
  - Video + speckle pattern [Zalevsky et al. 2009]



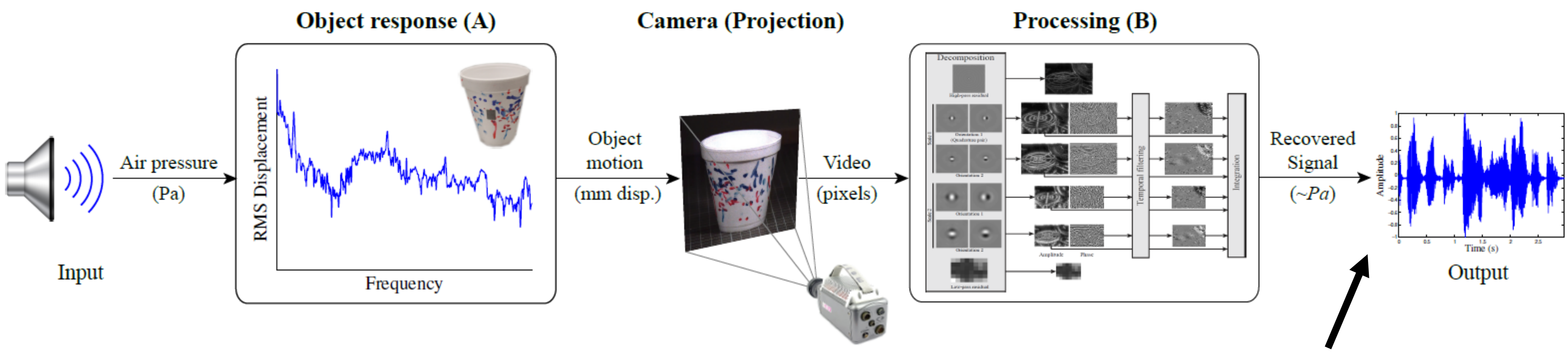
The Visual Microphone

- ~~Active~~ ***Passive*** technique to recover sound



The Visual Microphone

Physical
(unrelated to our processing)



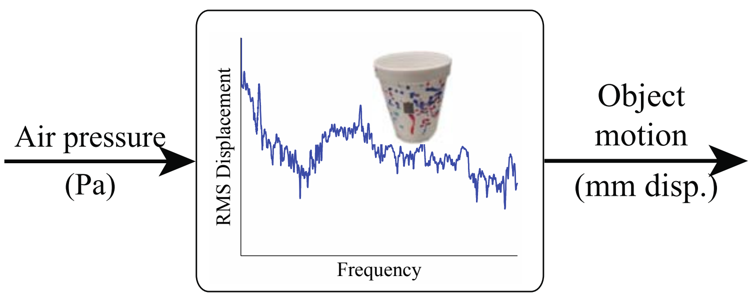
Unit-less
(but correlated with the
input pressure waves)

Assumption: Camera and object are static (any motion is due to sound)

Physical Analysis

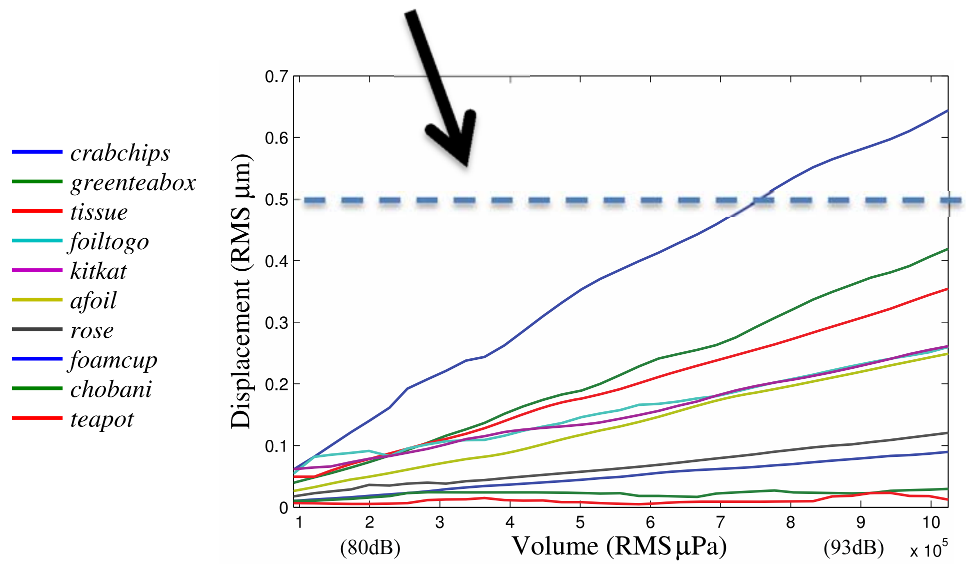
crabchips *greenteabox* *tissue* *foiltogo* *kitkat* *afoil* *rose* *foamcup* *chobani* *teapot*

Analysis



Object response (A)

0.5 RMS micrometers!

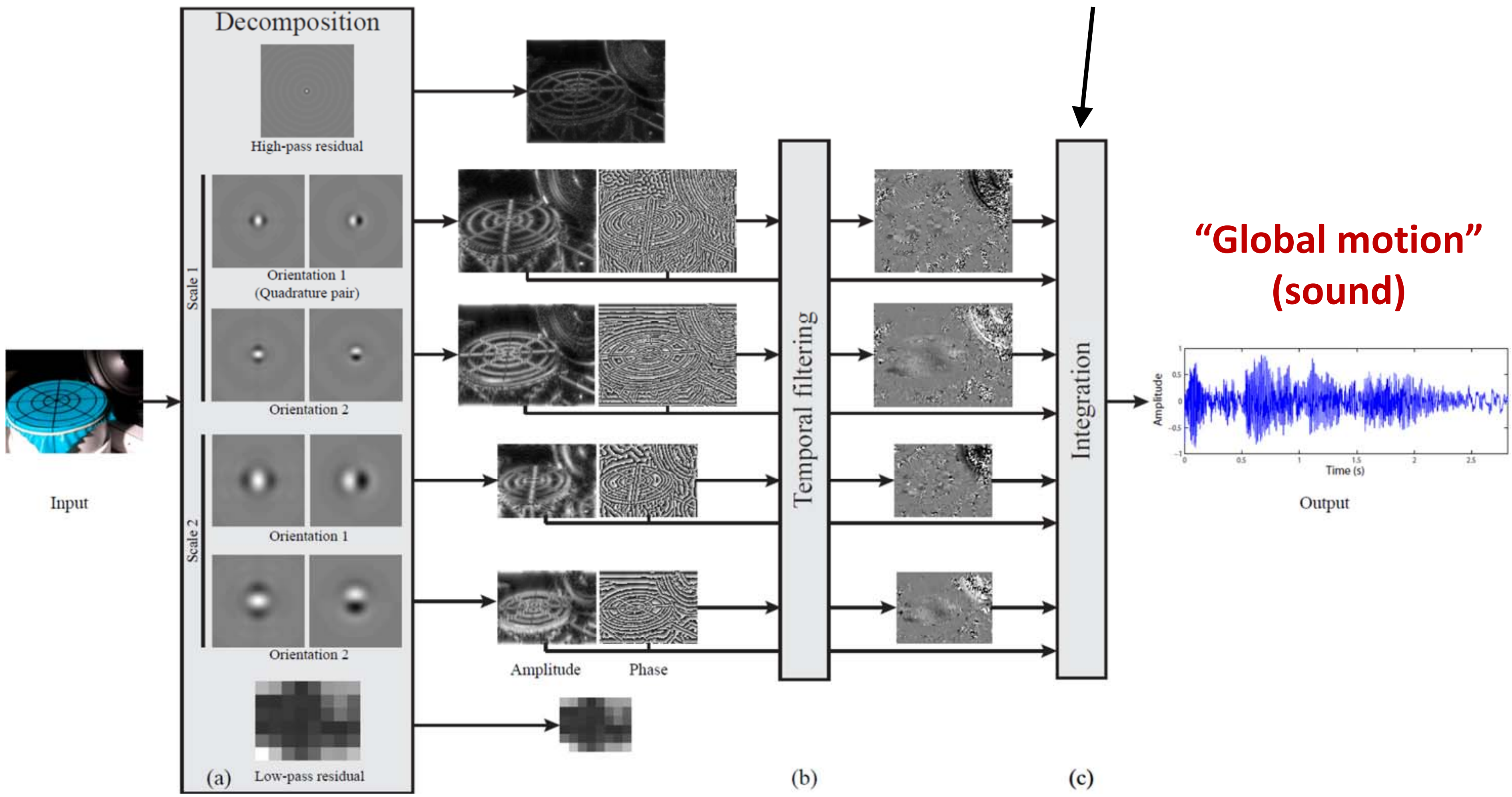


(b) Motion vs. sound volume

300 Hz at **90 dB** → **0.1 micrometers motion** (0.0001 millimeters!) → **<0.01 pixel displacement**

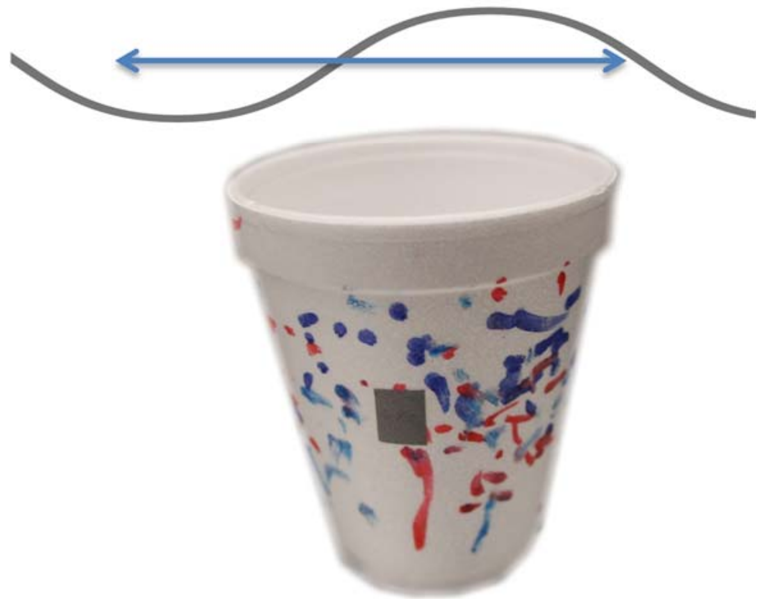
Video to Audio

Average local motions



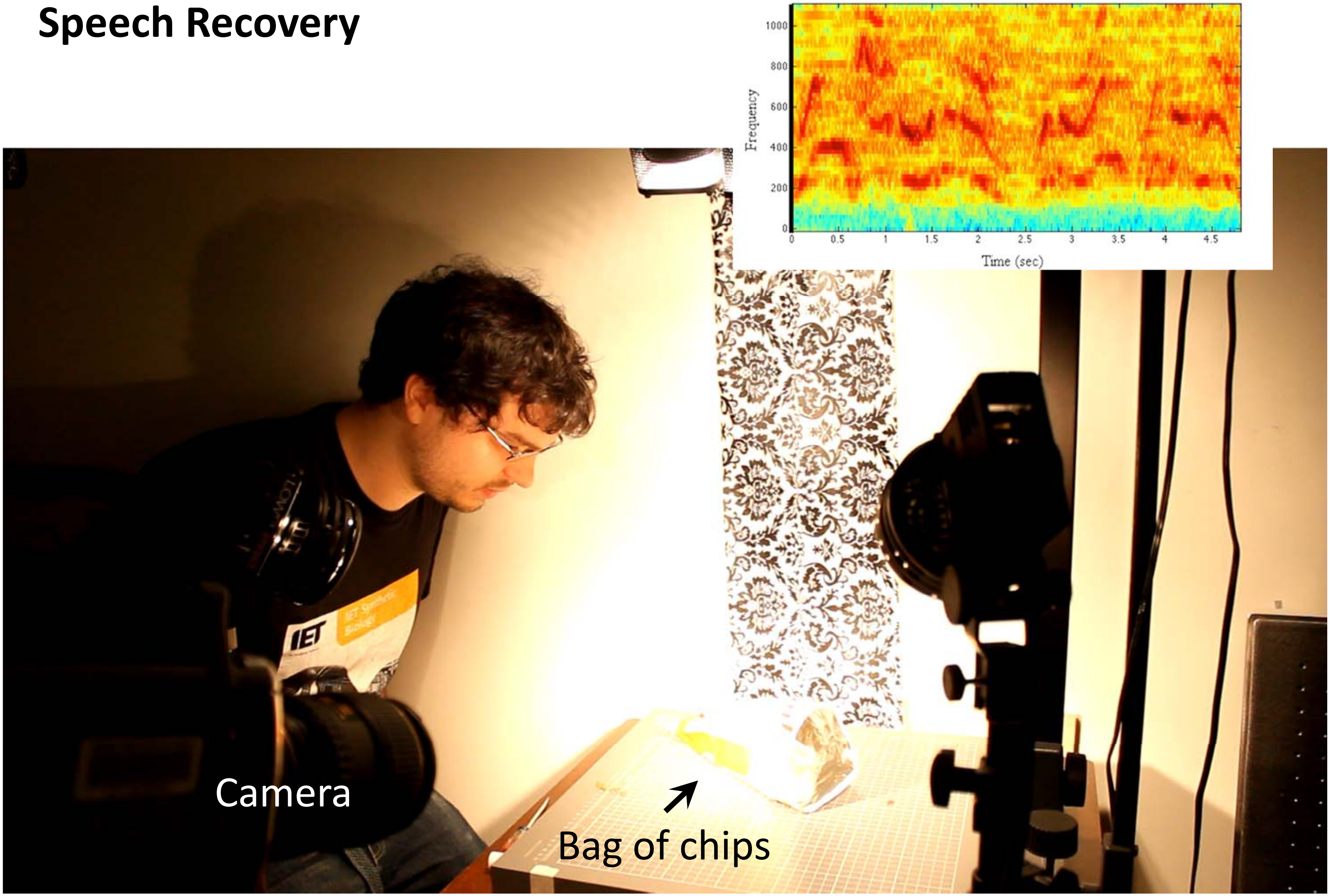
“Global motion” (sound)

Combining Local Motions

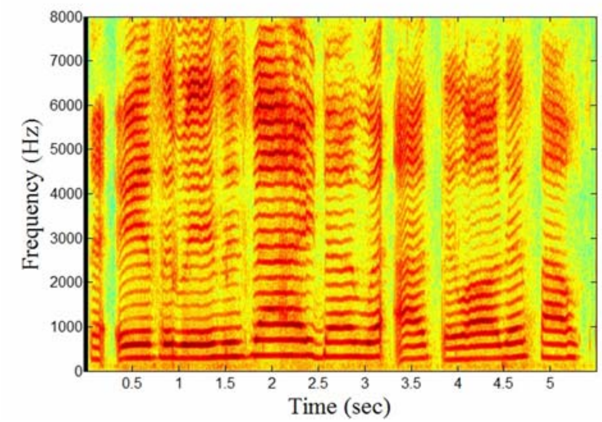


- Telephone frequencies
  - 300Hz to 3.4kHz
- Speed of sound in air
  - ~343 m/s

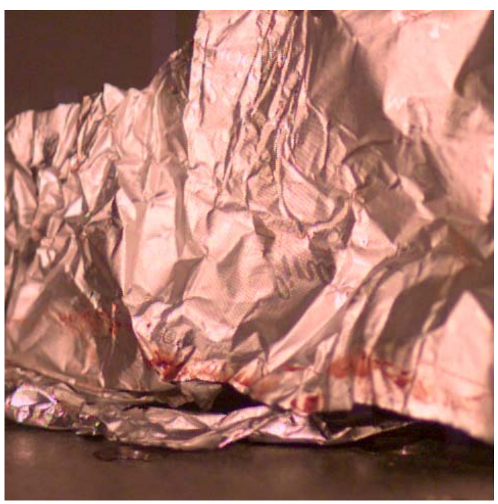
Speech Recovery

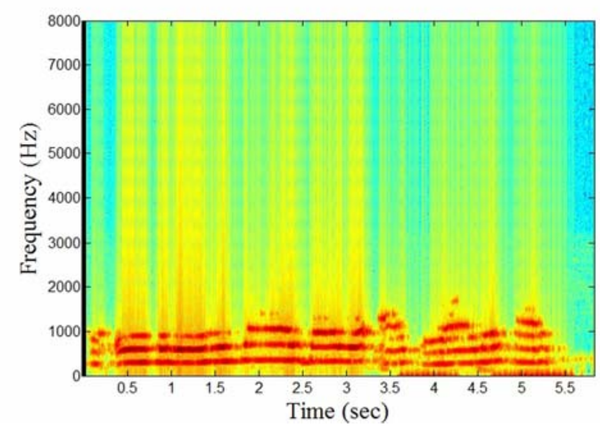


Itsy Bitsy Spider



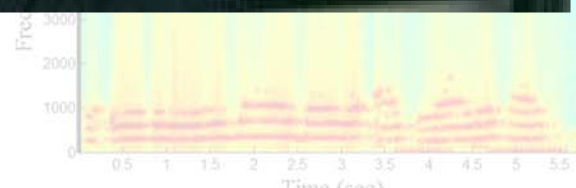
Candy wrapper (6420 fps)



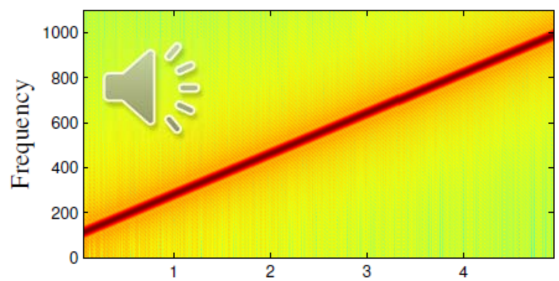


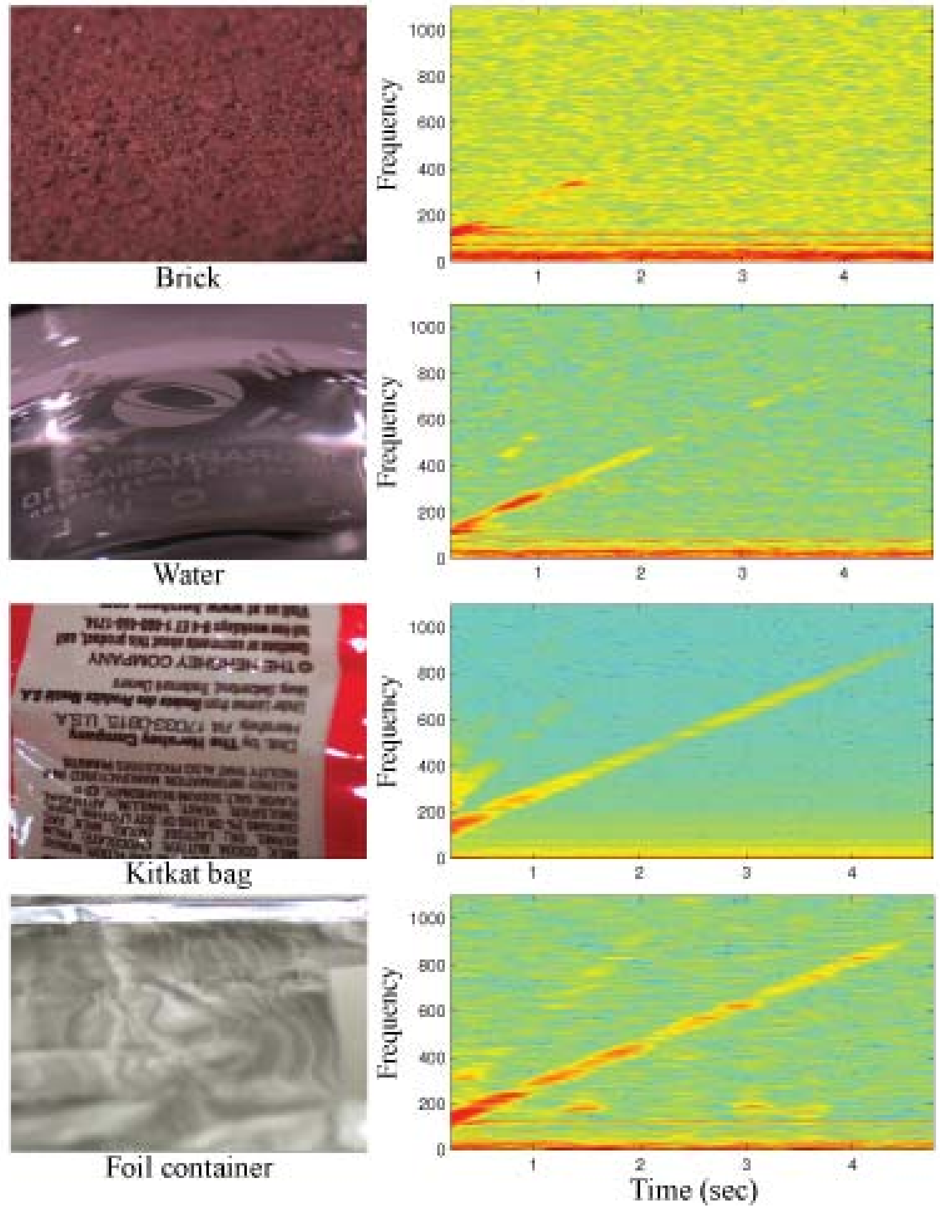
Itsy Bitsy Spider

Candy wrap



Testing Visual Microphones

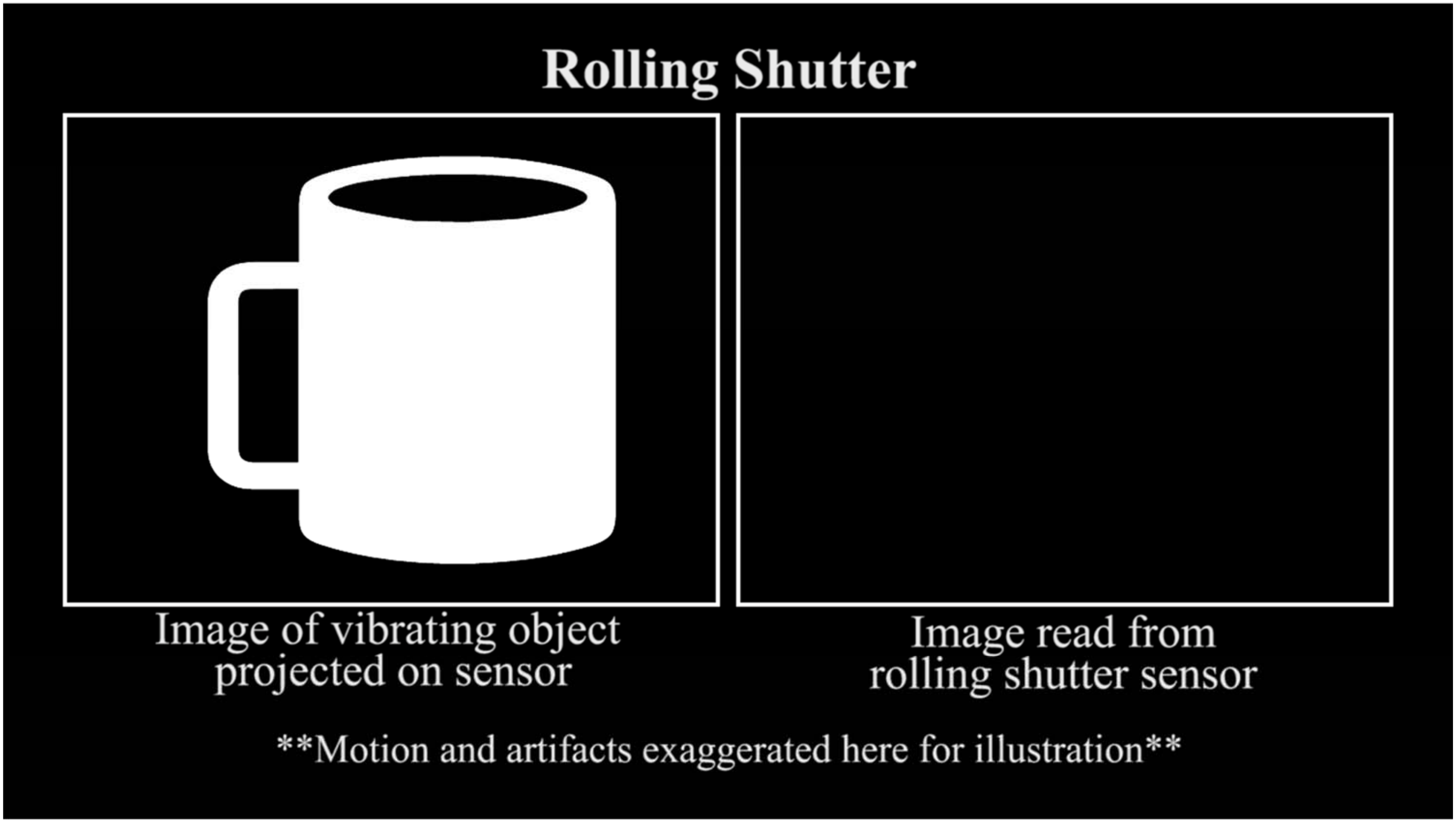




Rolling Shutter

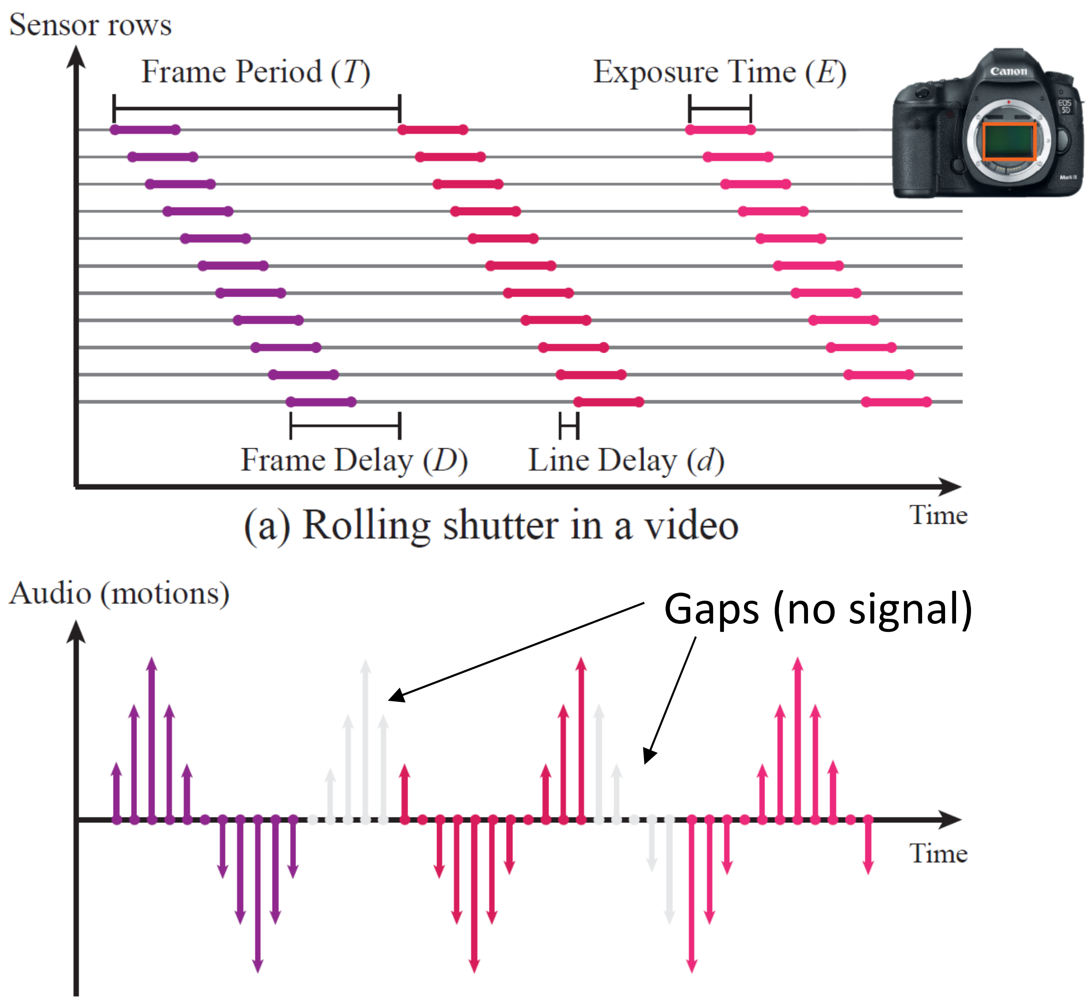
https://www.flickr.com/photos/sorenragsdale/3904937619/
http://www.flickr.com/photos/boo66/5730668979/

Rolling Shutter



Rolling Shutter

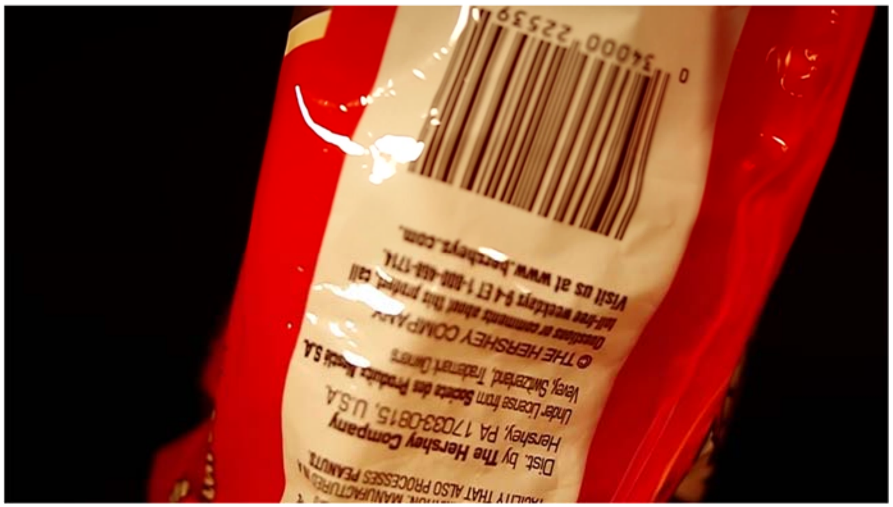
- Slow 2D camera → Fast 1D camera!



(a) Rolling shutter in a video

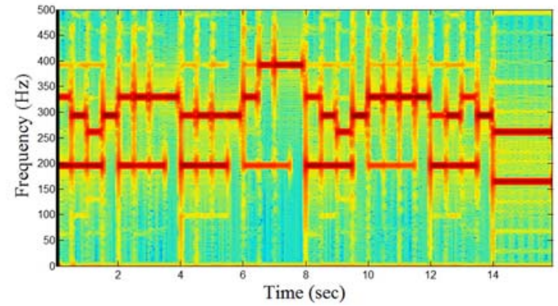
(b) Converted to audio signal

Rolling Shutter

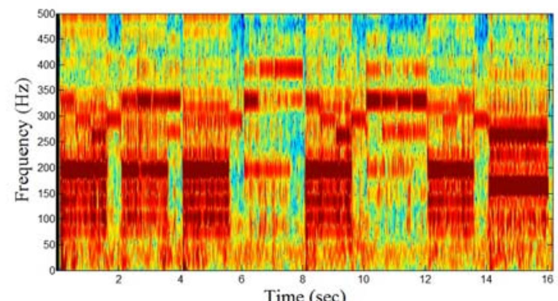


Input video (60 fps)

<u>Regular SLR</u>

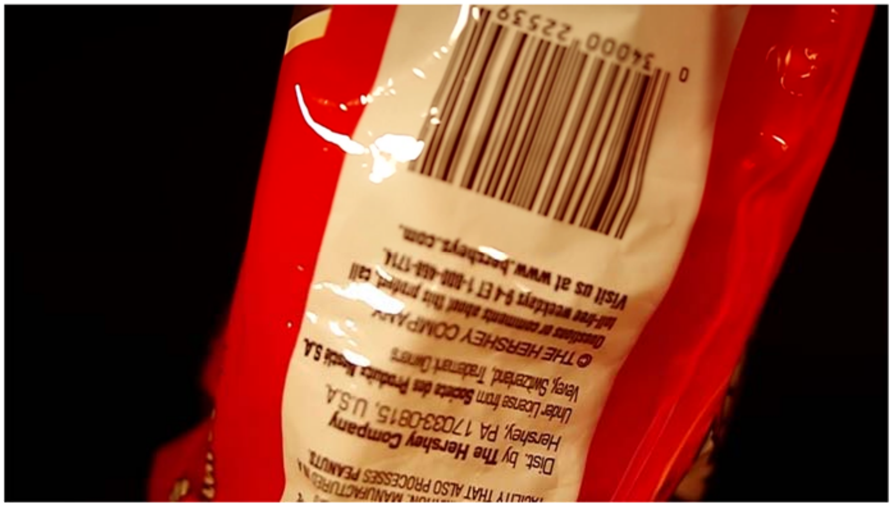


Input

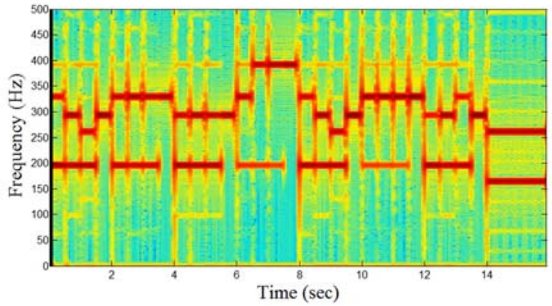


Recovered Sound

Rolling Shutter

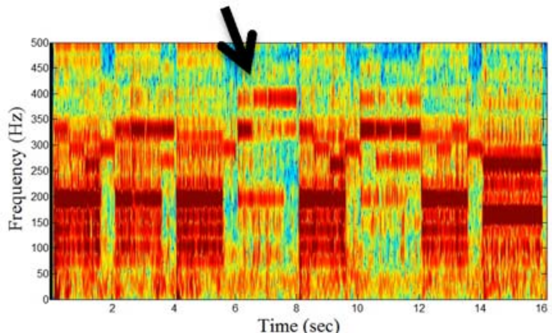


Input video (60 fps)



Input

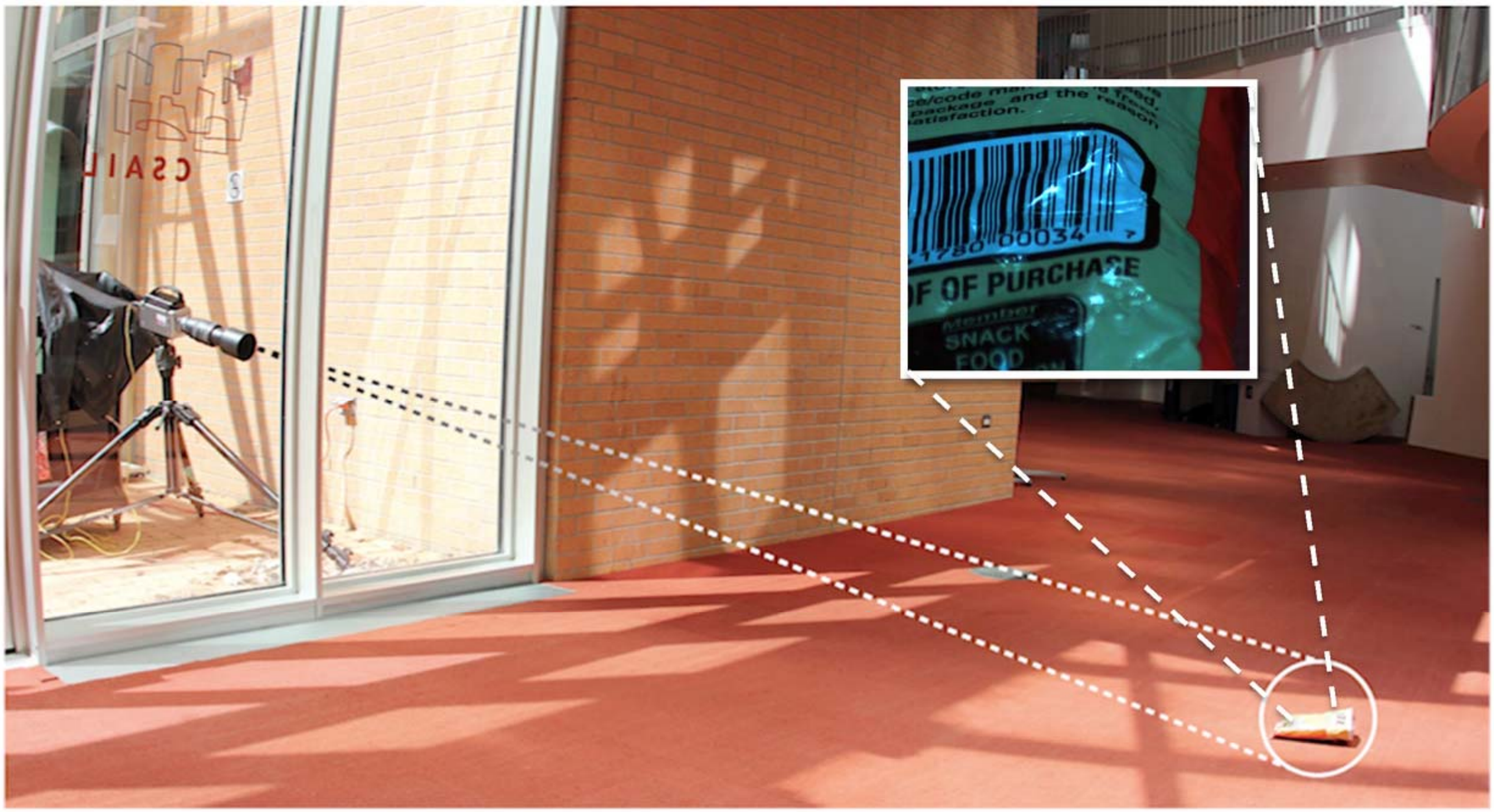
400Hz (>6 times the frame rate!)



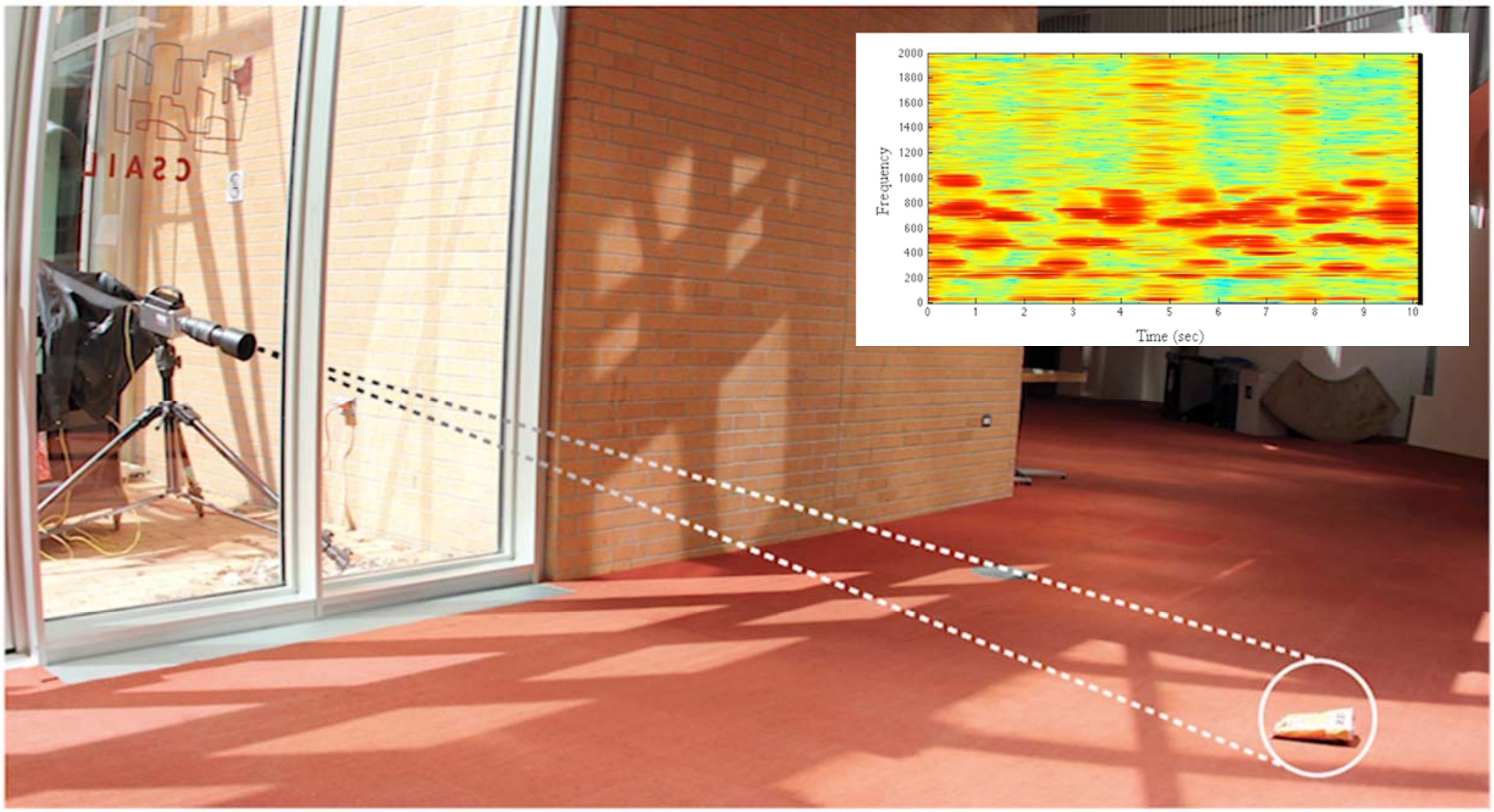
Recovered Sound

Range of Operation

4kHz, 400 x 480 video

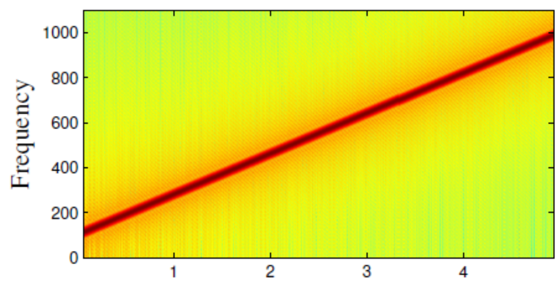


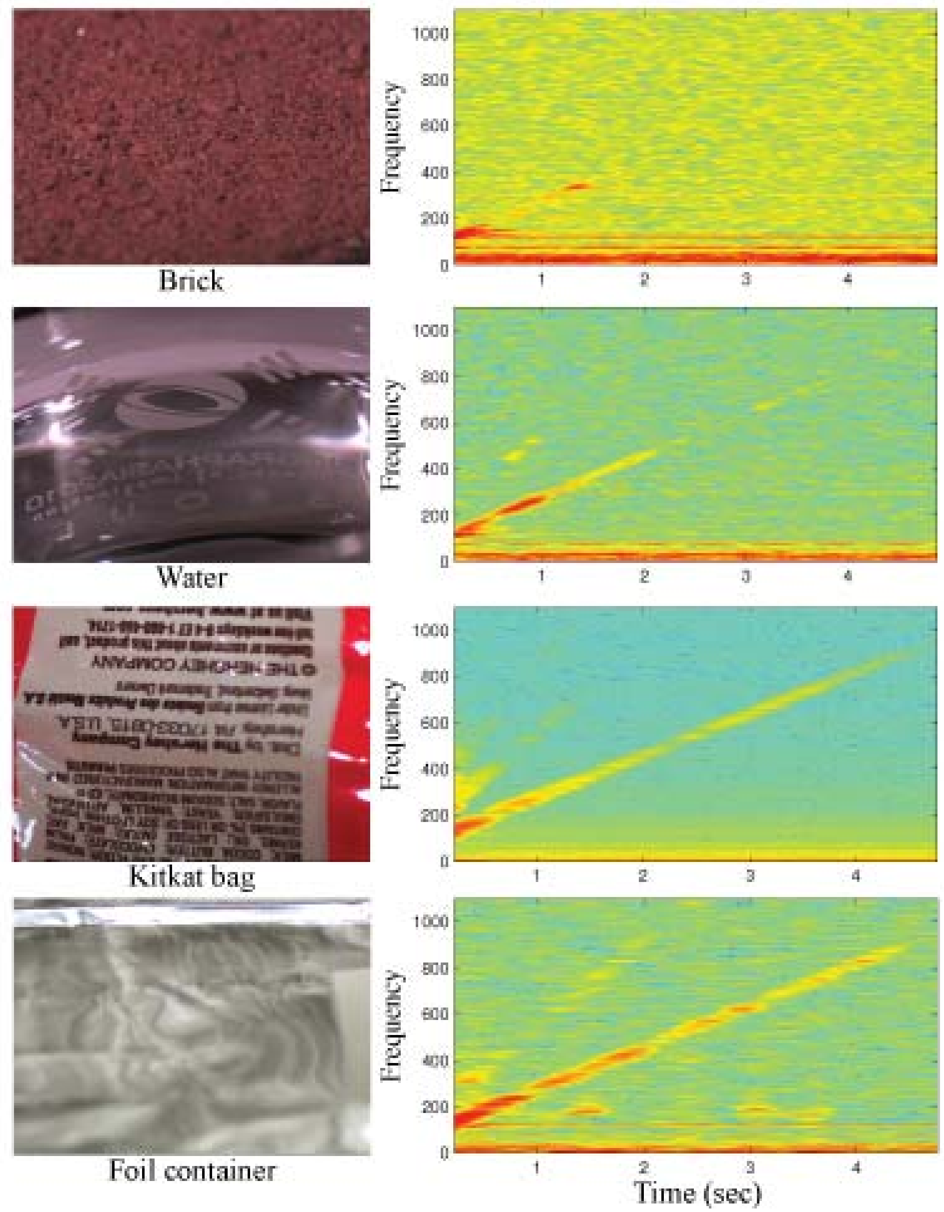
Range of Operation



Wait Wait.. Don't Tell Me! (Aug 30 2014)

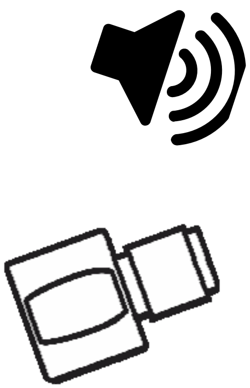
Vibration Depends on Object Properties

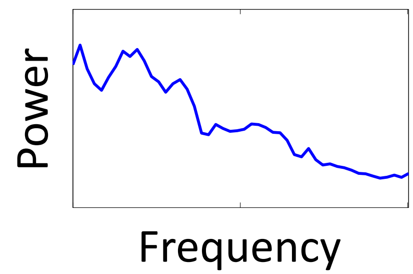




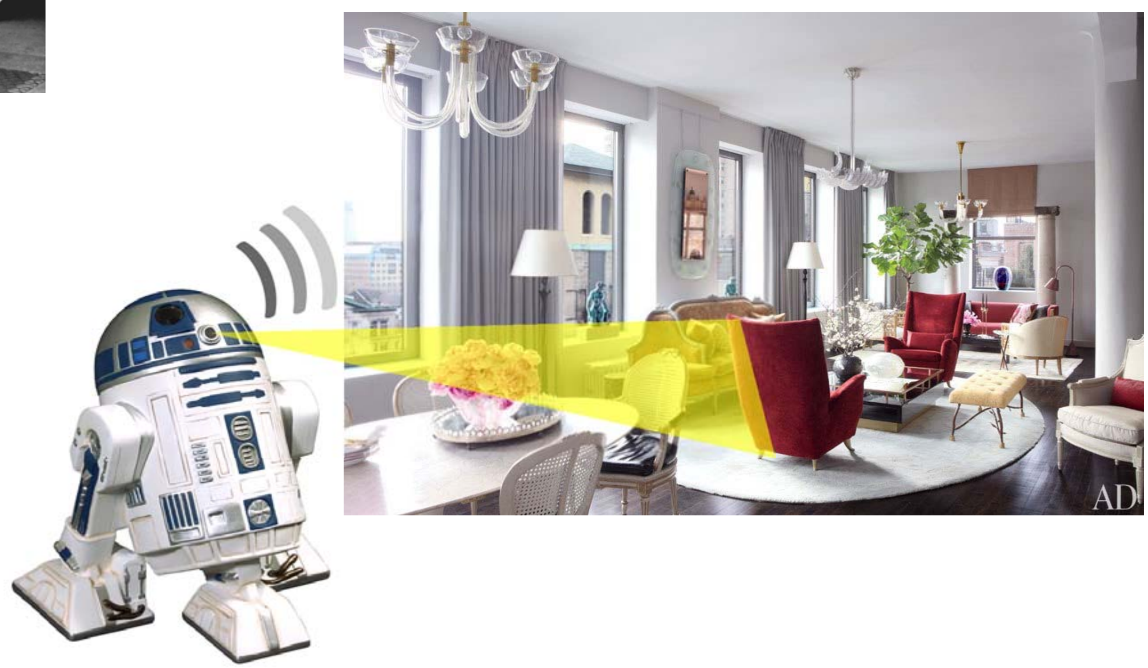
Estimating Material Properties with Sound and Cameras

Estimating Material Properties
from Small Motions in Video, CVPR 2015



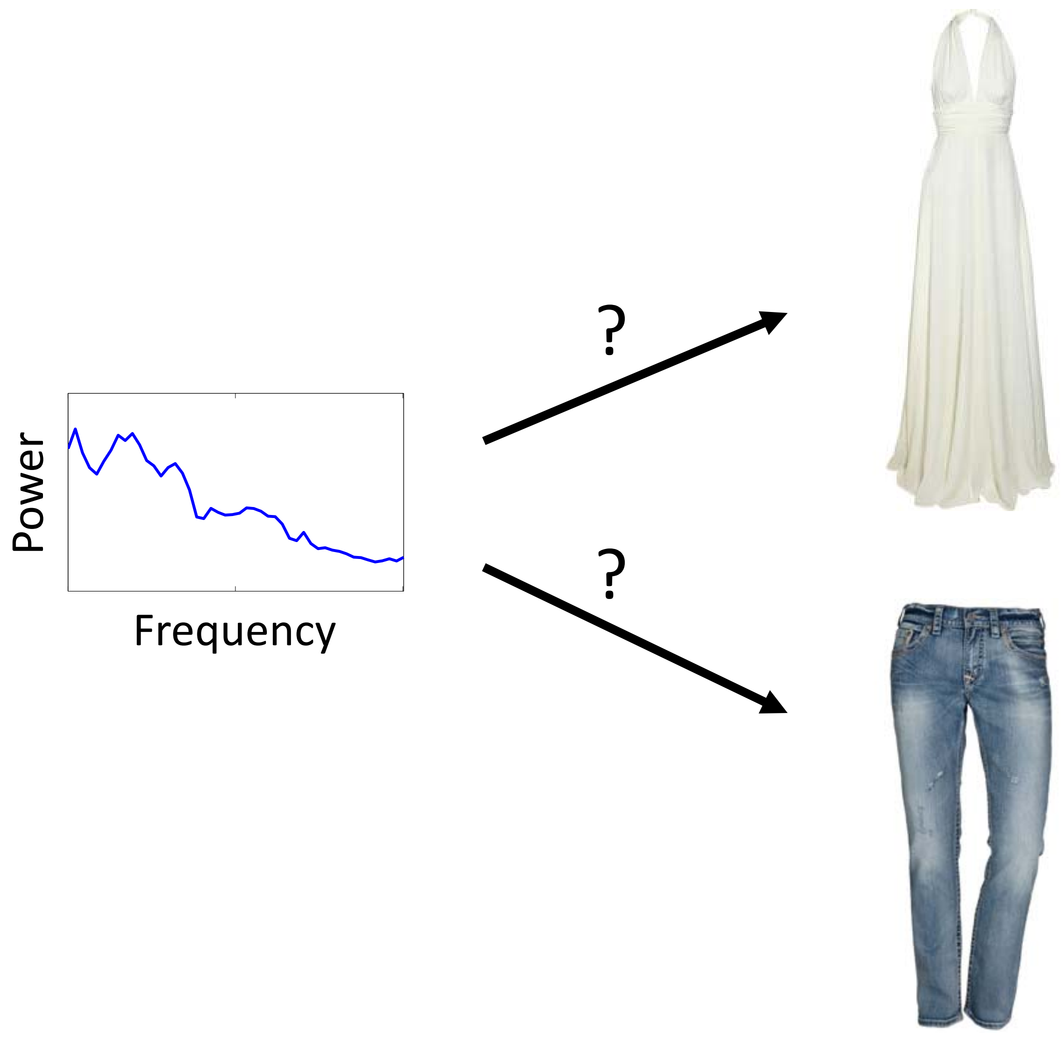


Material properties
(e.g. stiffness)



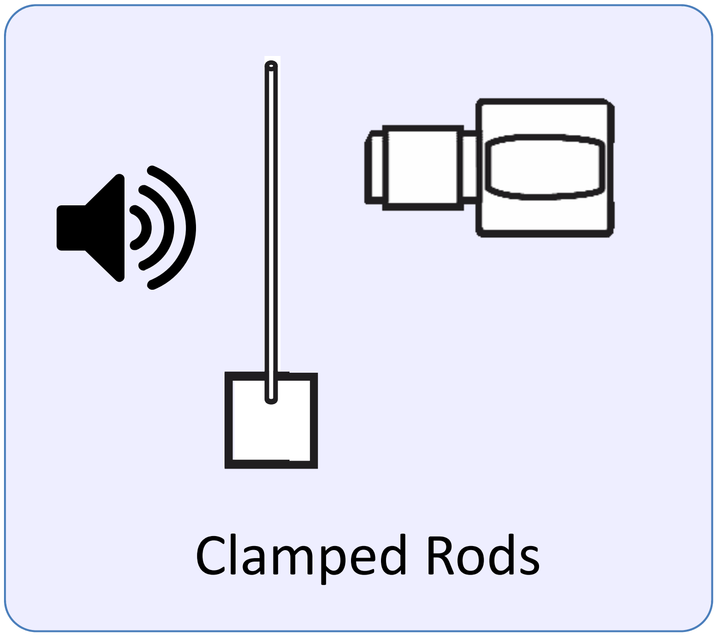
Problem: Hard to Disambiguate Geometry and Material

- Motion spectra reflects a combination of material properties AND shape



Experiments

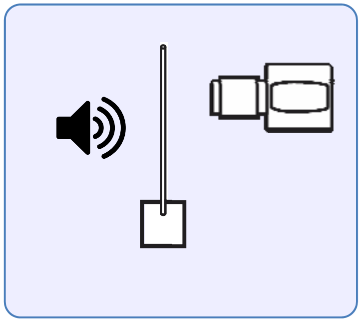
Estimating Material Properties from Small Motions in Video
Davis, Bouman, et al., CVPR 2015



Known Geometry

Unknown but
Similar Geometry

Material Properties in Clamped Rods



Diameter

Fundamental freq.

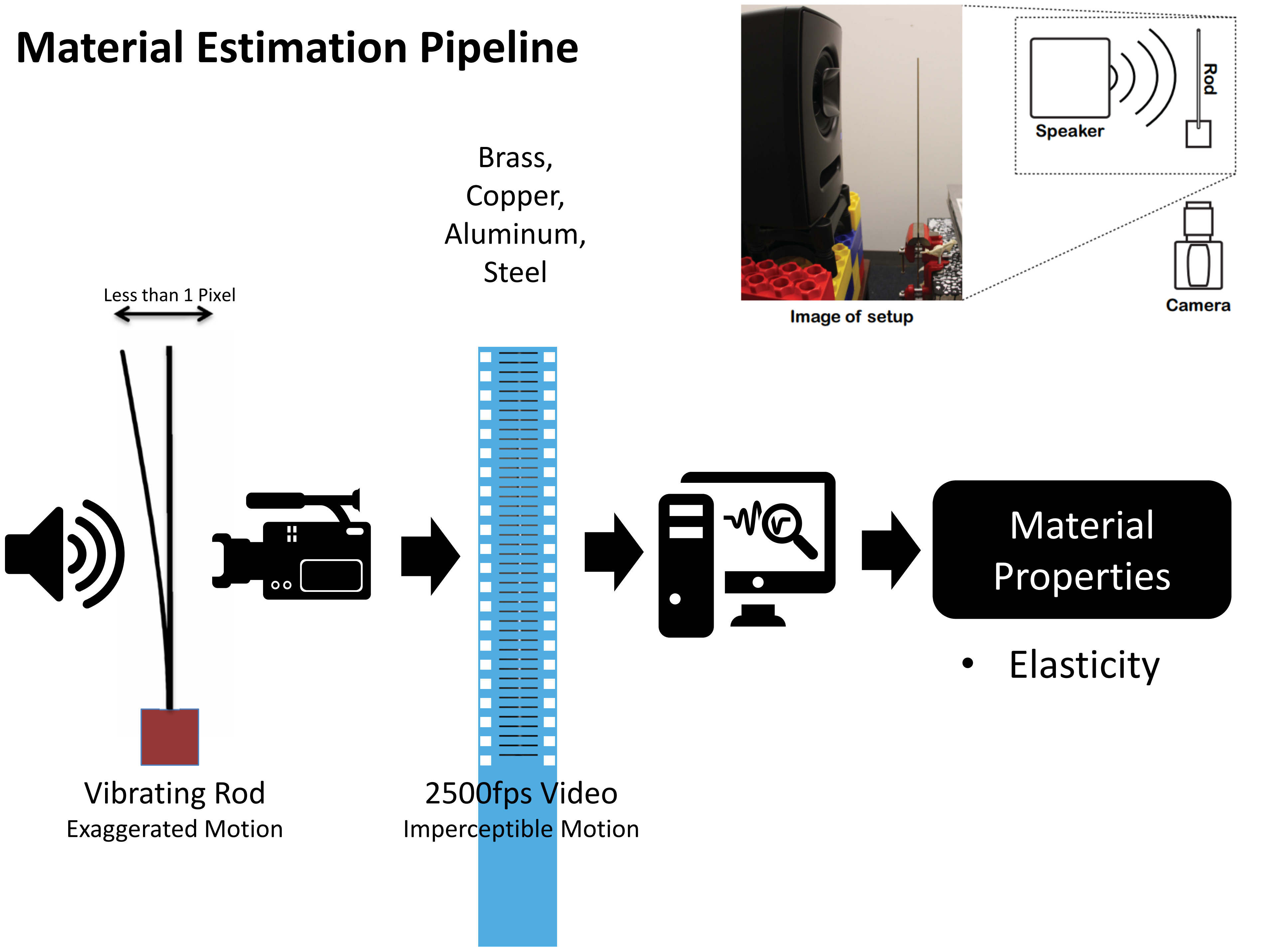
$$\omega_1 = 0.1399 \frac{d}{L^2} \sqrt{\frac{E}{\rho}}$$

Elasticity

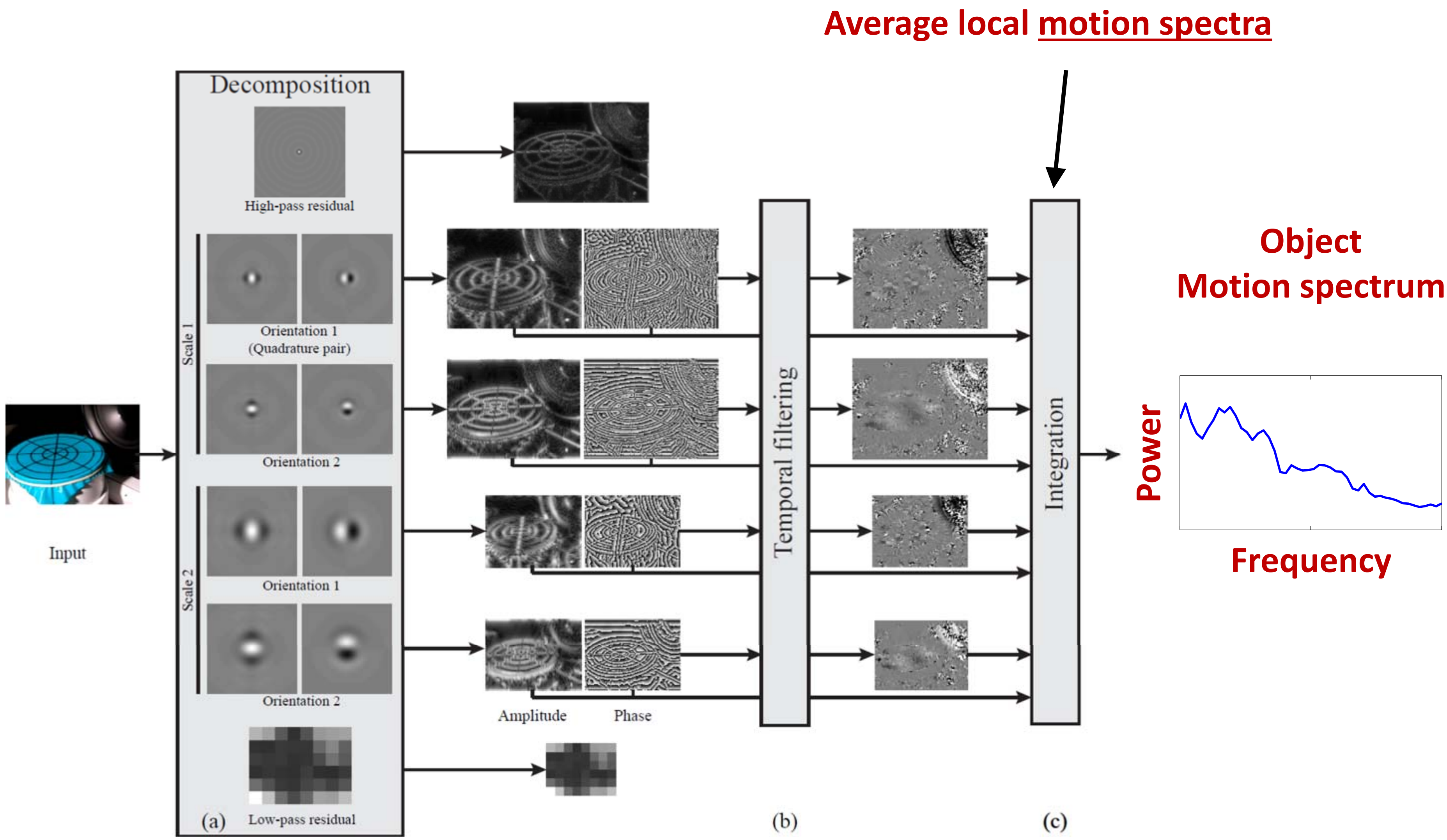
Density

Length

Material Estimation Pipeline

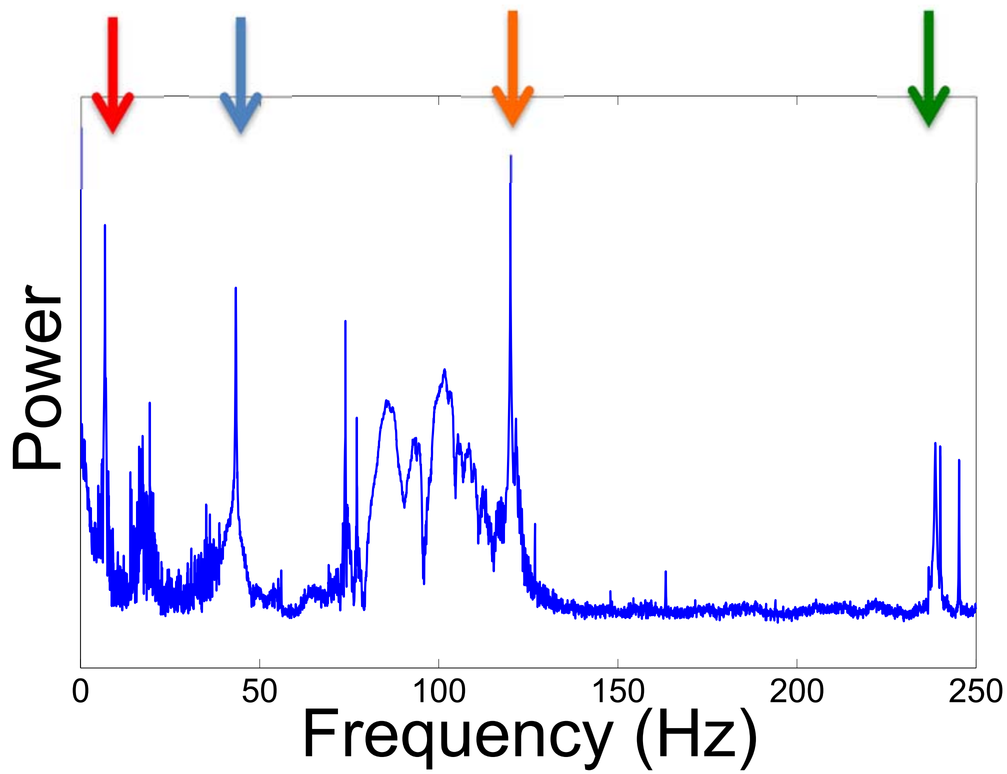


Processing



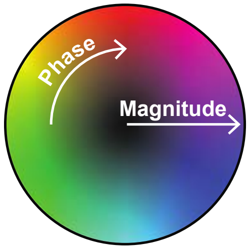
Average local motion spectra
Decomposition
High-pass residual
Scale 1
Orientation 1
(Quadrature pair)
Orientation 2
Scale 2
Orientation 1
Orientation 2
Low-pass residual
Input
Amplitude
Phase
Temporal filtering
Integration
Object
Motion spectrum
Power
Frequency
(a)
(b)
(c)

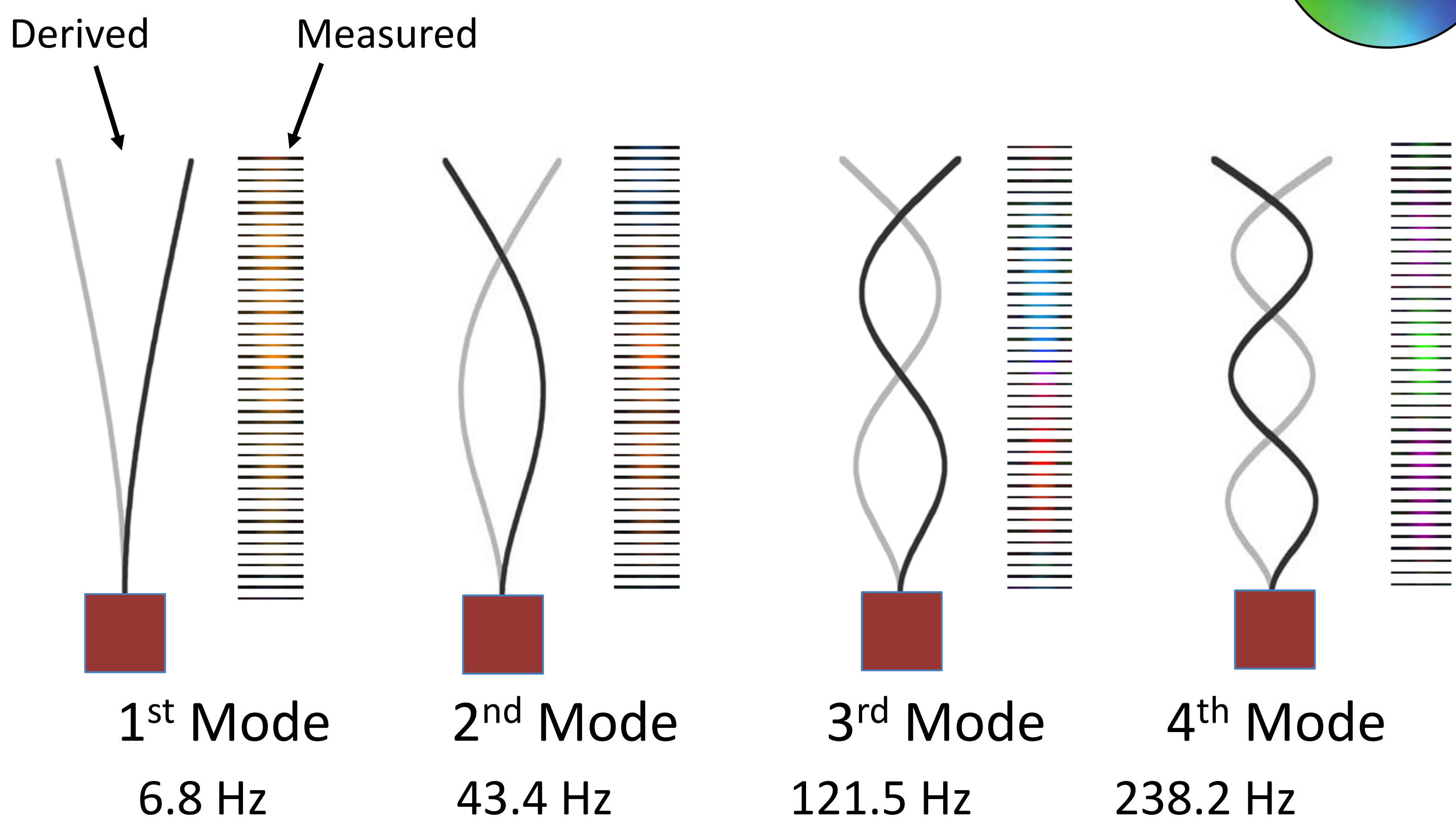
Recovered Resonant Frequencies



1st Frequency : 6.8 Hz
2nd Frequency : 43.4 Hz
3rd Frequency : 121.5 Hz
4th Frequency : 238.2 Hz

Verifying Recovered Modes: Phase Visualization





Material Properties from Recovered Frequencies

2 points for each rod, for two lengths we tired:
15 and 22 inches

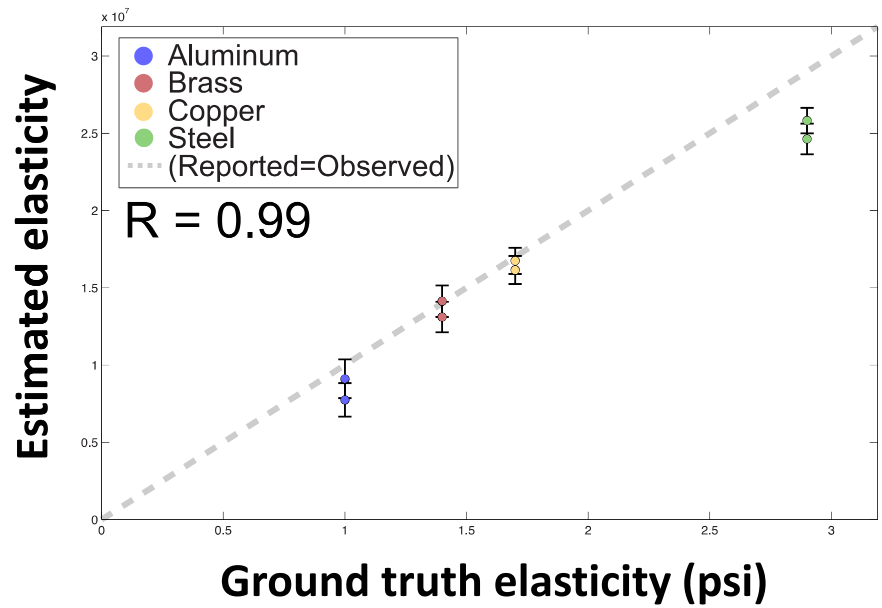
Diameter

Elasticity

$$\omega_1 = 0.1399 \frac{d}{L^2} \sqrt{\frac{E}{\rho}}$$

Density

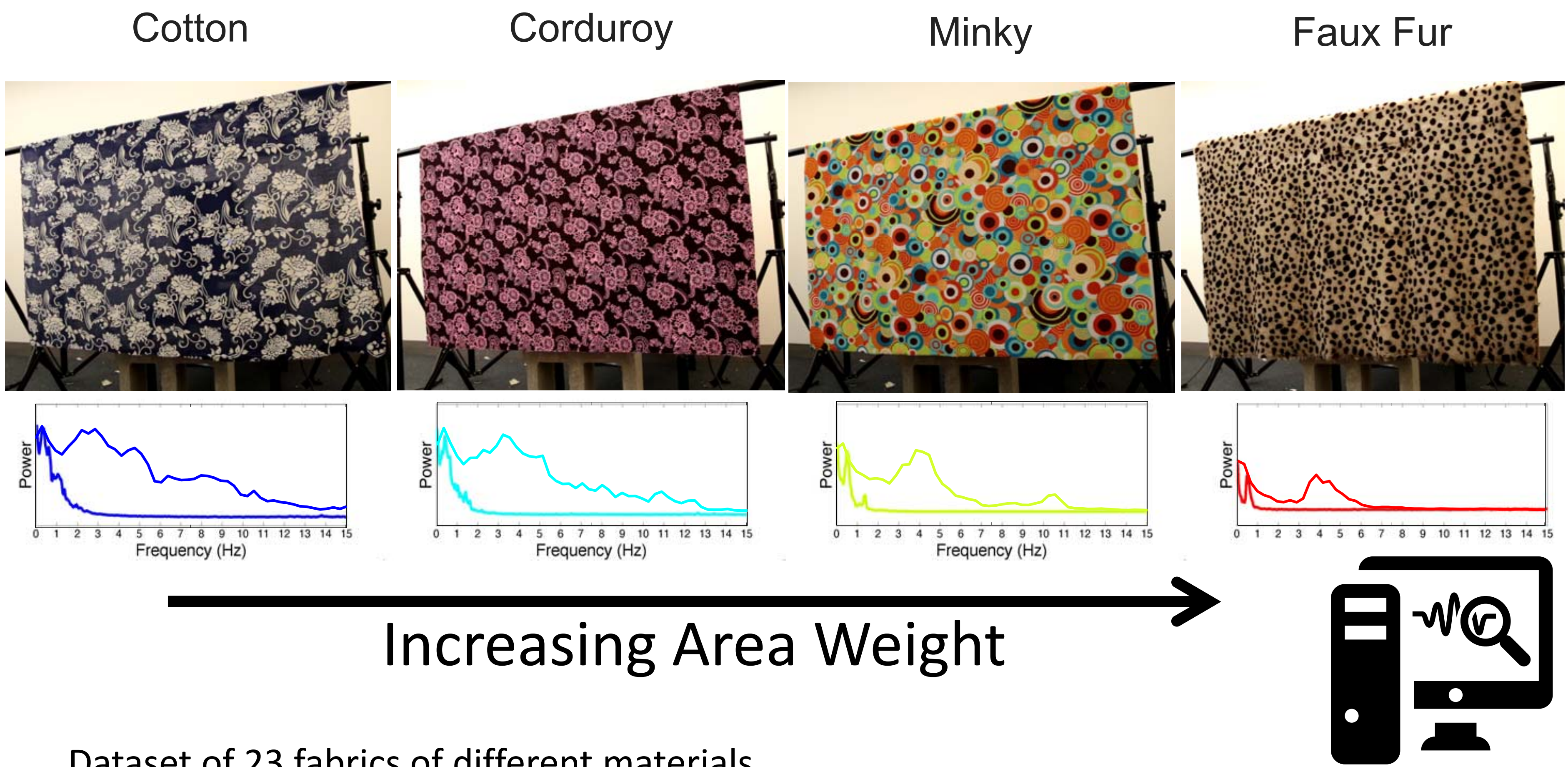
Length



Compute diameter, length, density (mass/volume) with ruler and scale, solve for elasticity

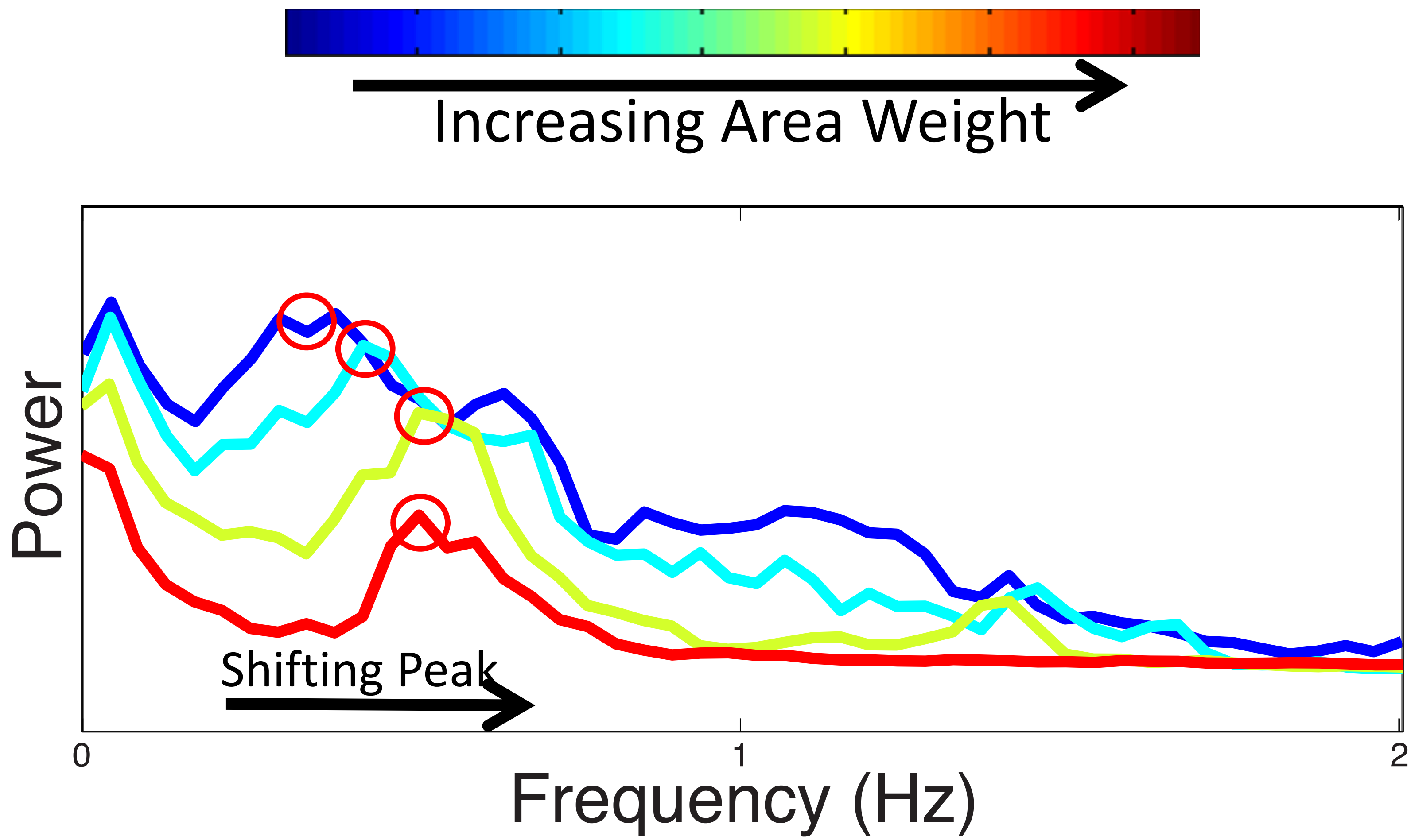
Good for Objects with Known (Simple) Geometry

Motion Signals in Fabrics

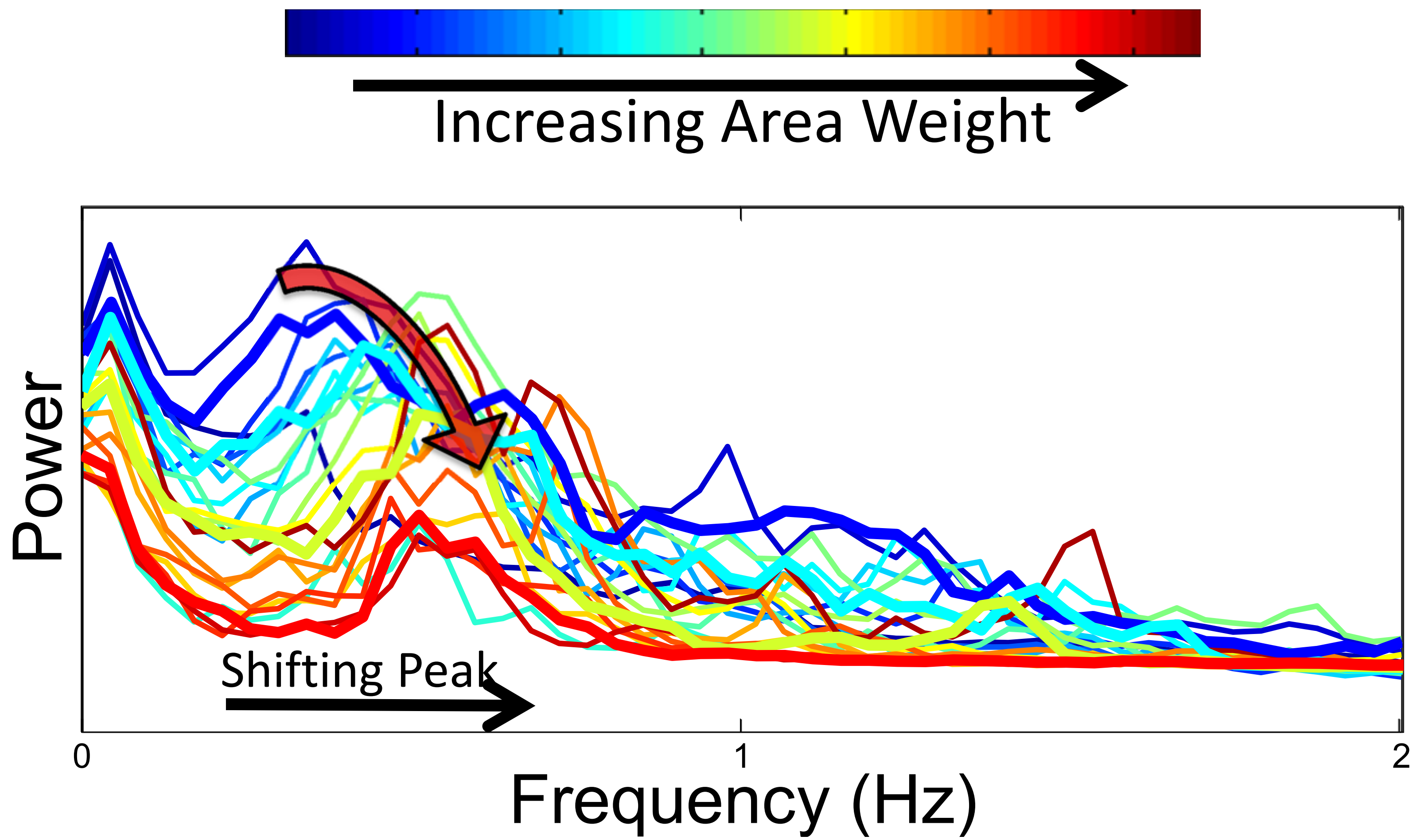


Dataset of 23 fabrics of different materials,
with lab measurements of area weight and stiffness

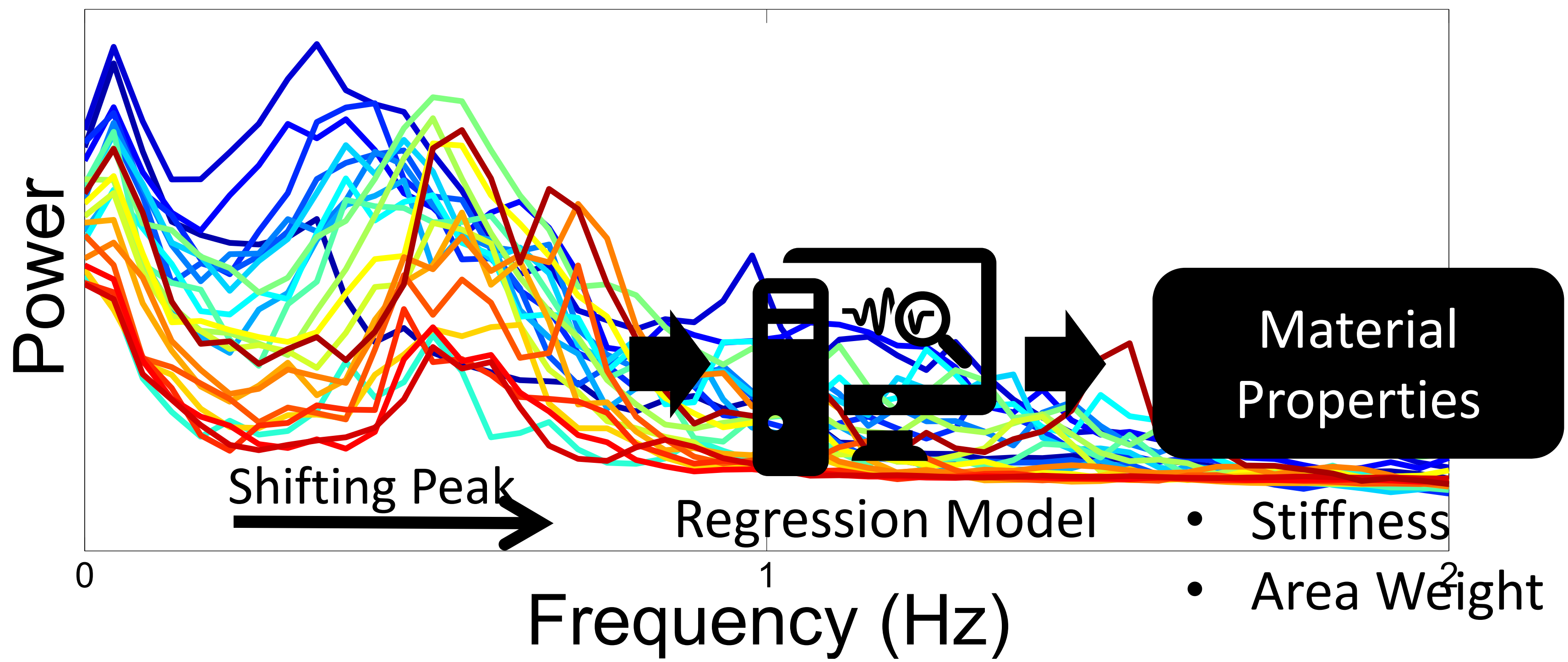
Identifying Trends in the Power Spectra



Identifying Trends in the Power Spectra



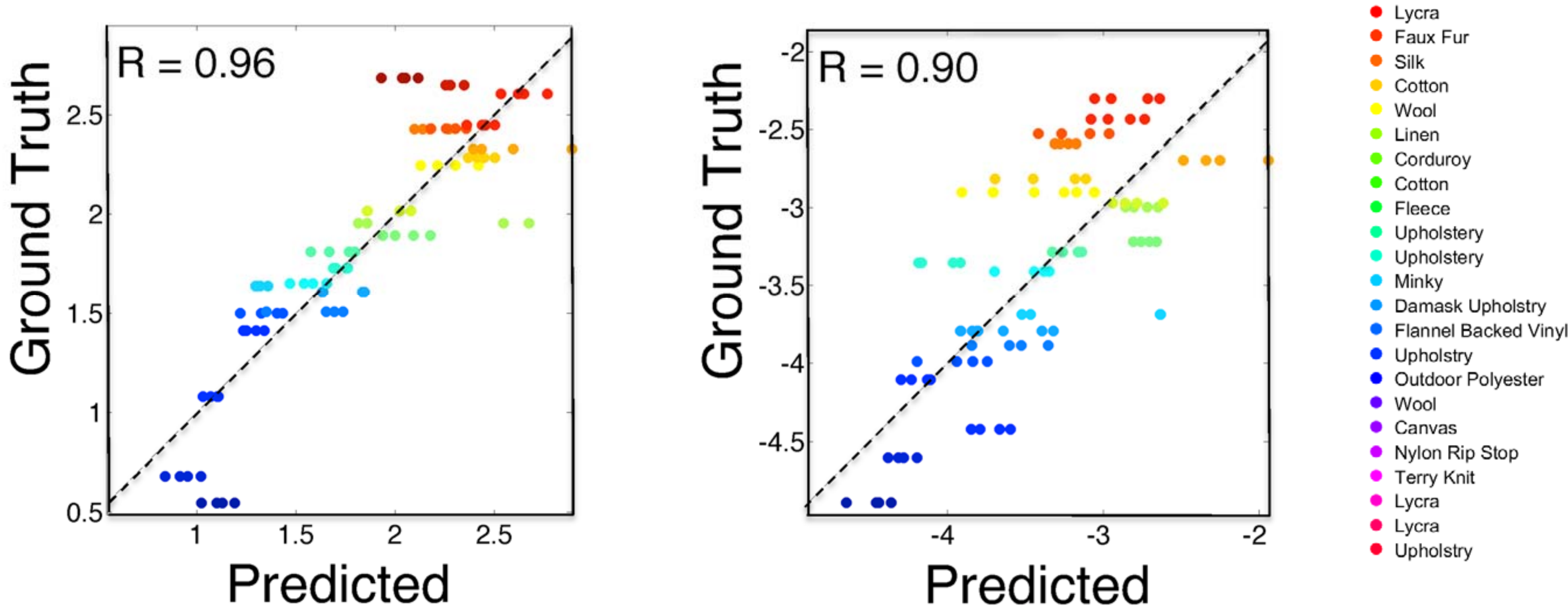
Learning Material Properties from Spectra



Estimated Fabric Properties

Area weight

Stiffness

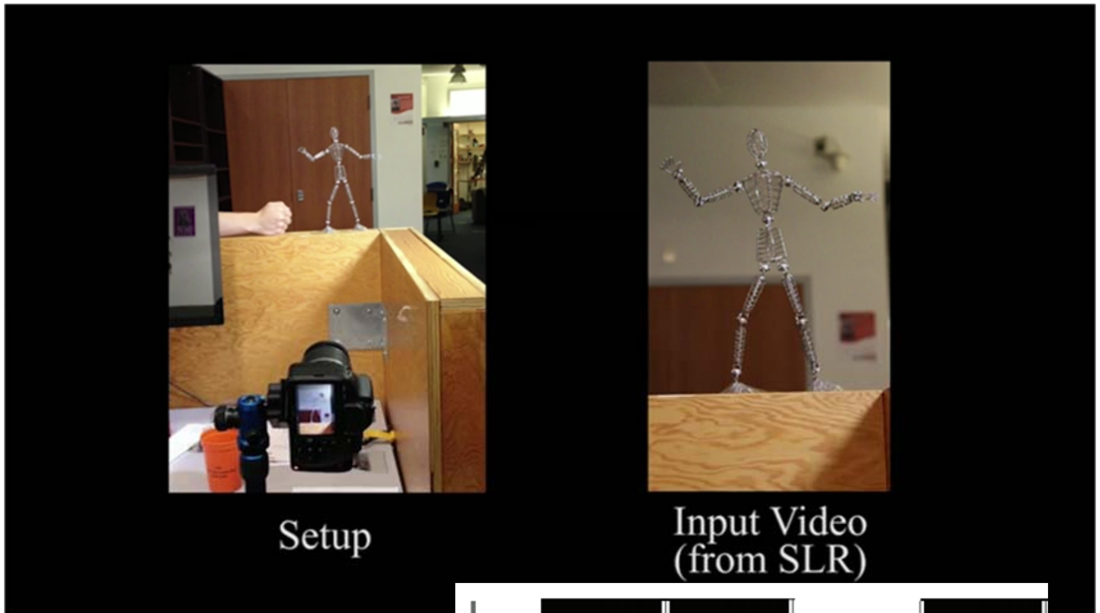


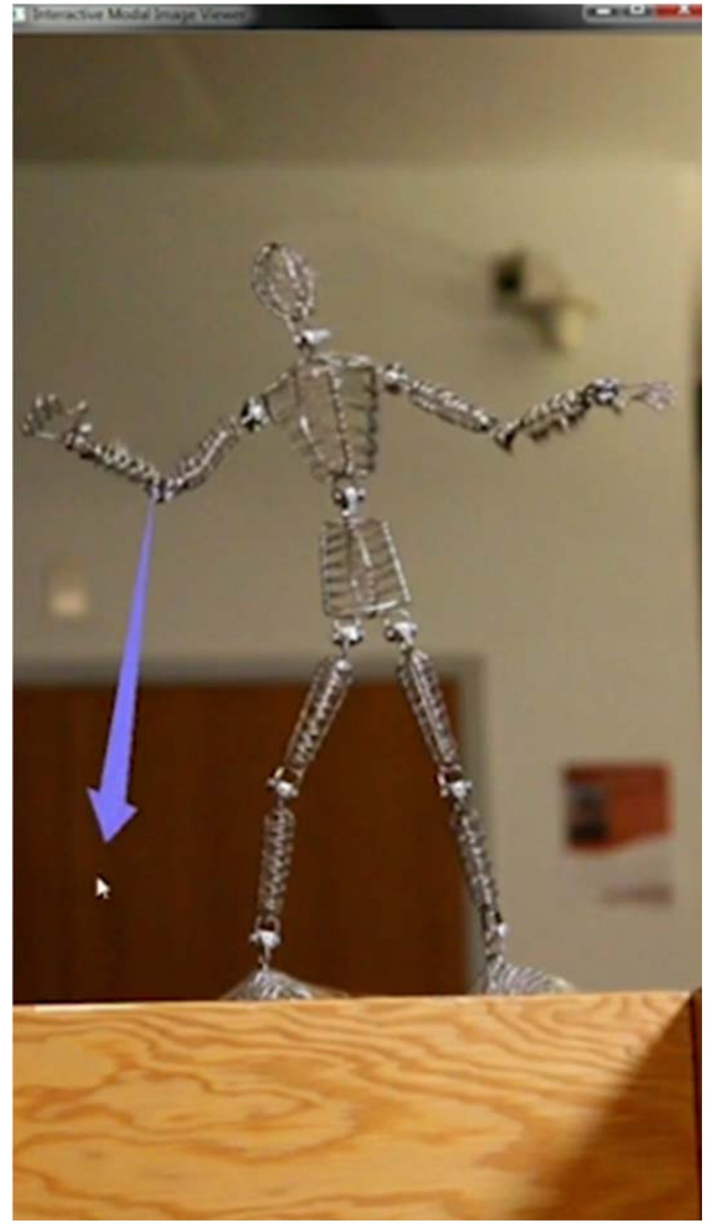
Ongoing Work

Abe Davis, Justin Chen, Fredo Durand

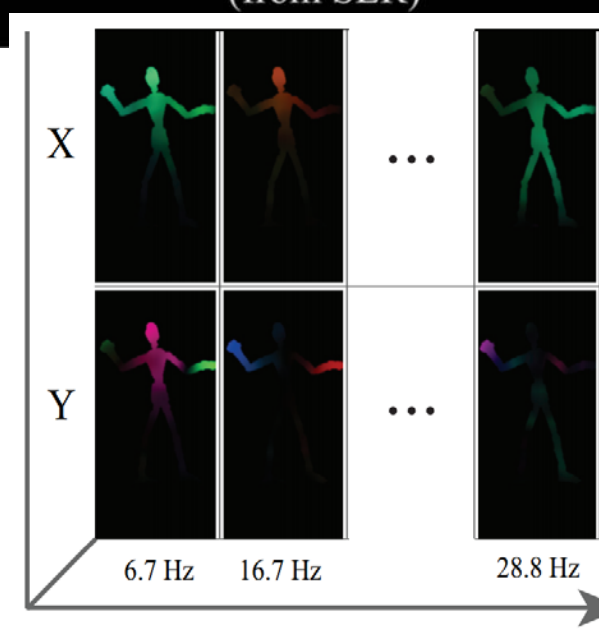
Image-Space Modal Bases for Plausible Manipulation of Objects in Video

SIGGRAPH Asia 2015



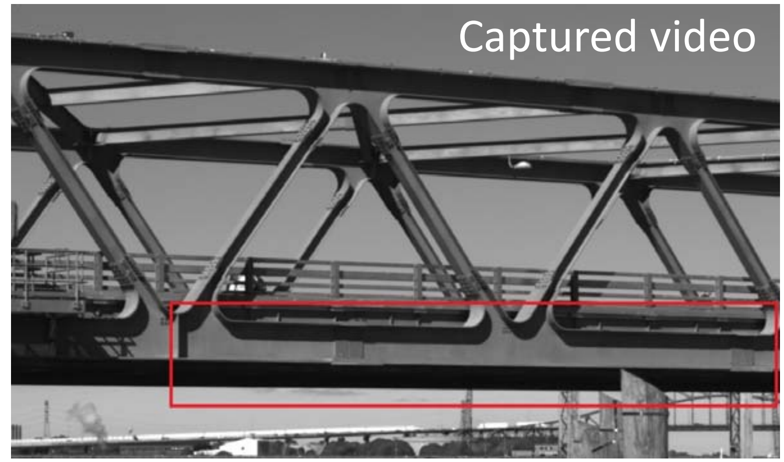


Extracted mode shapes



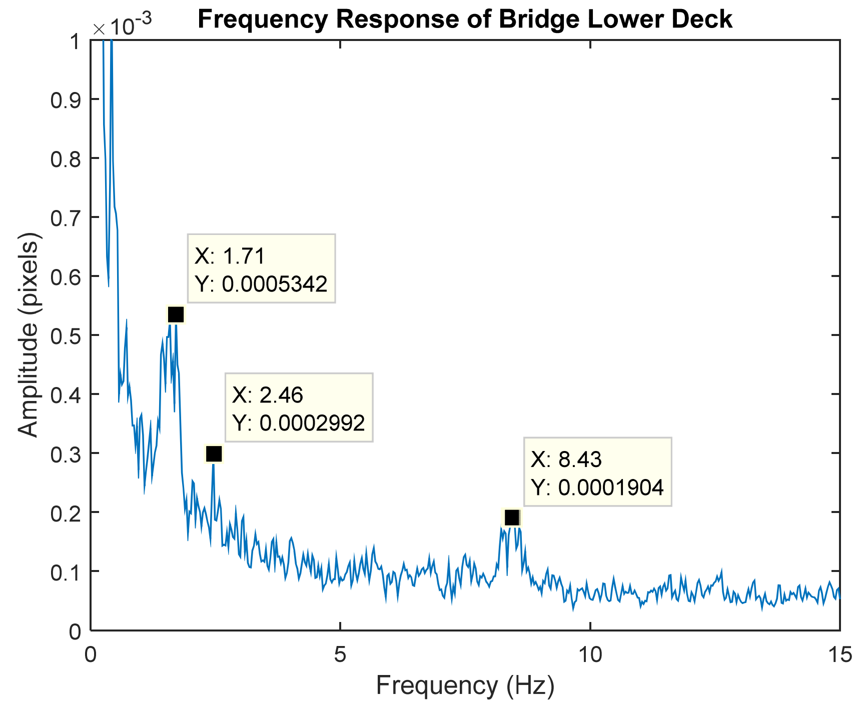
Ongoing Work

Justin Chen, MIT Civil Engineering,
with UNH and NHDOT
http://memorialbridgeproject.com/



Memorial Bridge, Portsmouth, NH

1184 x 700 pixels at 30 fps
~80 meters from bridge

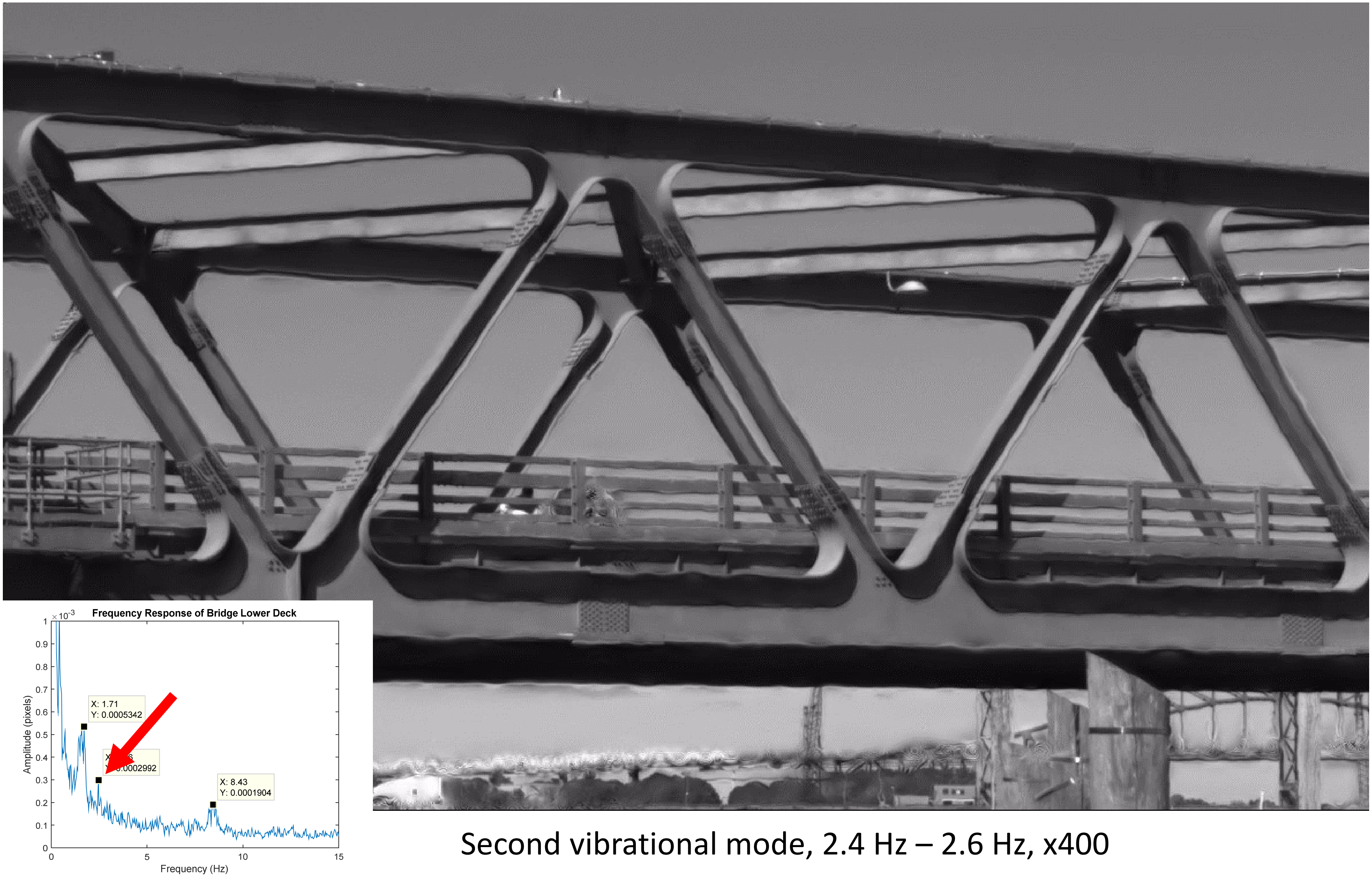


Ongoing Work

By Justin Chen with UNH and NHDOT
http://memorialbridgeproject.com/

Ongoing Work

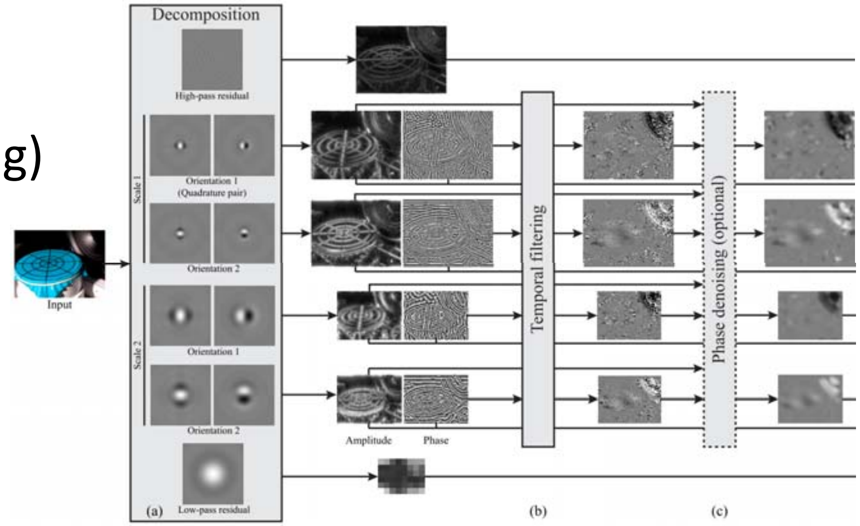
By Justin Chen with UNH and NHDOT
http://memorialbridgeproject.com/



Second vibrational mode, 2.4 Hz – 2.6 Hz, x400

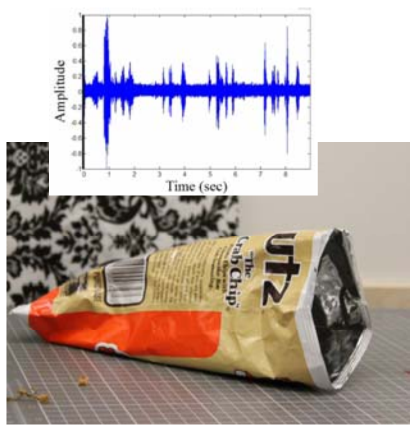
Summary: Visual Vibrometry

One basic pipeline
(phase-based processing)

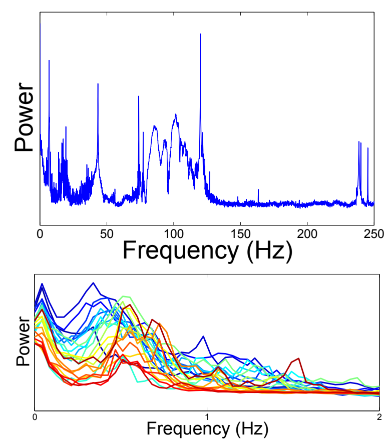


Magnification
(visualization)

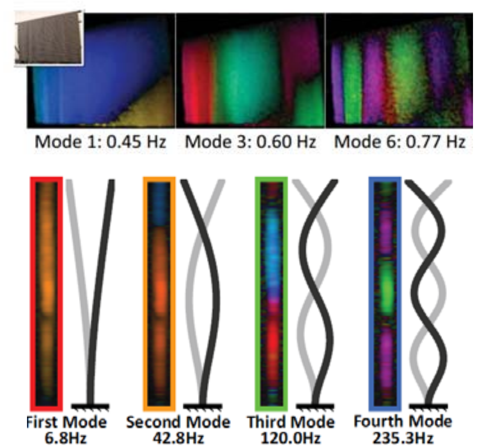
Sound recovery



Material property
estimation



Mode shape analysis



Miki Rubinstein, mrub@google.com
Google Research, Cambridge MA
(Work done while I was at MIT, Microsoft, Google)

Joint work with: **Bill Freeman, Fredo Durand, Neal Wadhwa, Abe Davis, Katie Bouman, Justin Chen, Gautham Mysore**

Project pages, code, demos:

Video magnification portal: http://people.csail.mit.edu/mrub/vidmag/

Phase-based processing: http://people.csail.mit.edu/nwadhwa/phase-video/

The Visual Microphone: http://people.csail.mit.edu/mrub/VisualMic/

Material properties: http://visualvibrometry.com

TED talks by Abe Davis, Miki Rubinstein