**Sound representation of picture**

- Sound representation of picture
- Object presented on screen.
- Scrubbing finger across screen
- RGB value of touch point.
- By function to represent picture by
- Social interactions.
- Further aid these people to ease
- Social interactions.
- For second part we also build a
- Function to represent picture by
- Synthesise sound to represent the
- RGB value of touch point. By
- Scrubbing finger across screen
- And listen to sound changes, one
- Can roughly tell the shape of
- Object presented on screen.

**Gender classification process**

- Camera capture
- Crop head and shoulder
- Face detection
- Feature extraction by CNN

**Results**

We trained the Logistical Regression with CNN feature vectors of
1662 pictures (1123 male, 579 female) that was downloaded from
ImageNet dataset and a stock photo website.

- From these pictures we cropped 2070x(1111 male, 959 female) head
- And shoulder regions by face detection and rotation with face angle.
- We also cropped all face regions from them.

- We experimented with different combination of training vector
- Count, vector length, also experimented training with full picture,
- Head and shoulder crop, and face crop, and found “head and
- Shoulder” crop gives best results. We have ~94% accuracy when
- Validating with training data.

- Then we tested our result with a new dataset (1090 pictures in total)
- That were not used in the training. Due to the randomness inherited
- From CNN, different test runs show different results.

- On average we have 68-78% accuracy of predicting gender correctly,
- Interestingly we have around 68% accuracy predicting back figure
- Without seeing face at all.

**Thanks!**

Professor Girod and Wetzstein
for a wonderful course.
TA Chen and Boin for helps on everything.

**References**

Compressing Deep Convolutional Networks Using Vector Quantization.
Imagenet classification with deep convolutional neural networks.