Using machine learning to enable voice & gesture interactions in educational web applications

Using machine learning to enable voice & gesture **interactions in educational web applications**

Why drag & drop interactions?

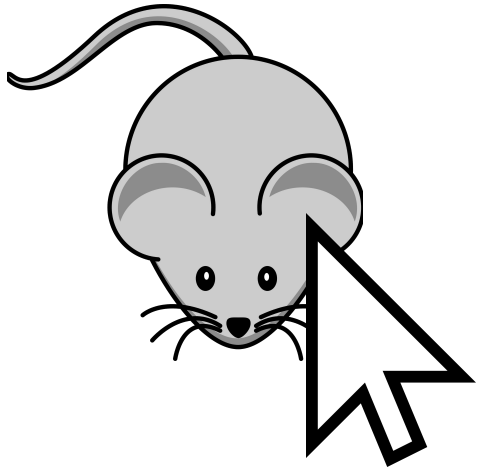
DEMO APP

Exploratory: No “correct” path

- Move an object to examine its effect
- Affordance = **immediate feedback**

Assessment: “Correct” path

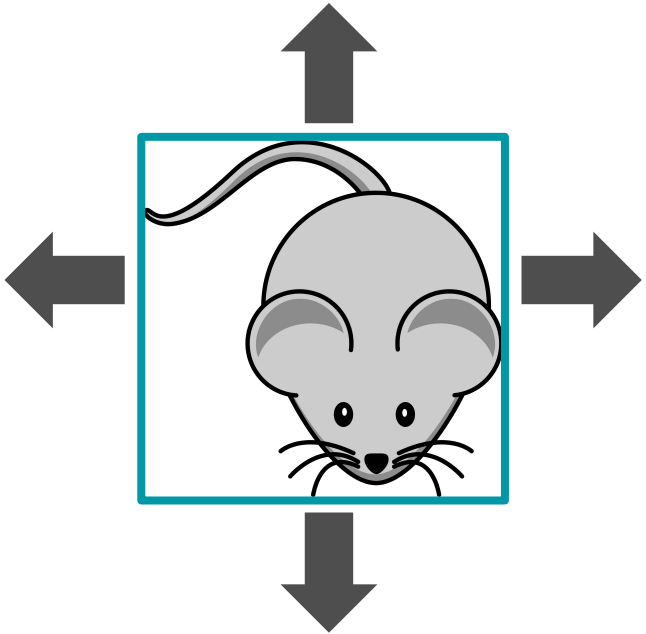
- Match, sort, rank, or label items
- Affordance = **automated grading**



😸 Using a mouse to move a mouse

Enable keys to operate

DEMO APP



What it does

- Enables Tab, Space, or Arrow keys to focus on object, pick up object, move object in increments or between targets, and drop object

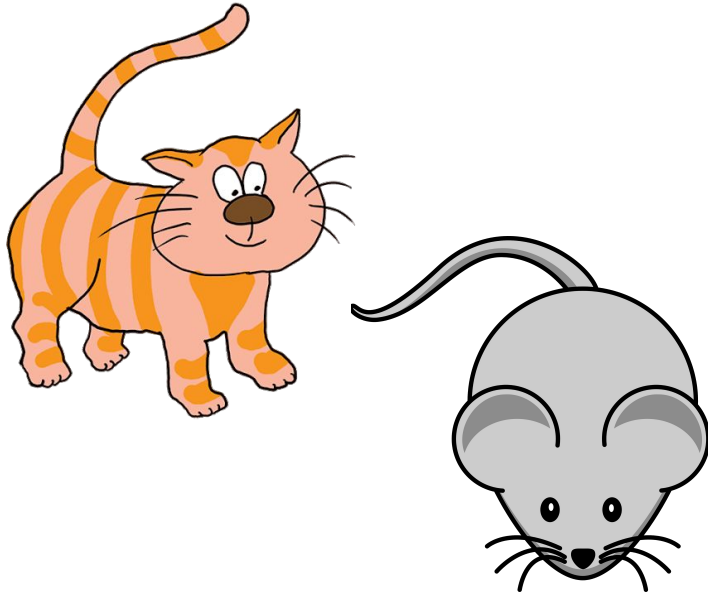
How it works

- Uses HTML5 native functionality + enhancements provided by our own custom elements

😸 Visible focus & No keyboard **traps** (mouse joke!)

Include screen reader alerts

DEMO APP



"*The cat is near the mouse*"

What it does

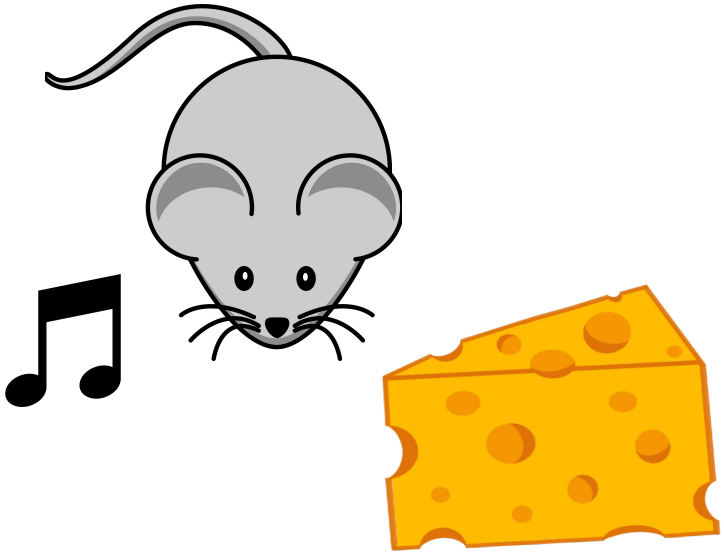
- Announces the status of the selected object

How it works

- Uses the React.js library created by Facebook to dynamically update ARIA live region

Include sound effects

DEMO APP



What it does

- Increases pitch or tempo as an object nears a target

How it works

- Uses Web Audio API to play sound
- Not supported on Internet Explorer

😸 Synthesized sounds can be **grating** (cheese joke!)

Using machine learning to enable voice & gesture interactions in educational web applications

What is machine learning?

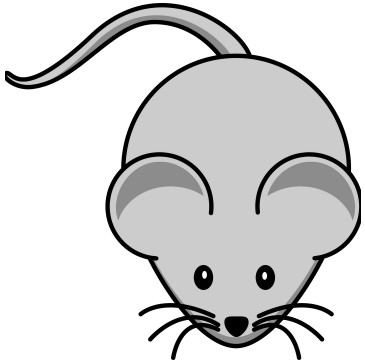
- Provides systems the ability to automatically learn to perform tasks and improve from experience without being explicitly programmed

- **Neural networks** are a form of machine learning which seek to replicate how the human brain works by creating layers of “neurons” that can be trained to recognize patterns in data

Sample of the MNIST dataset for handwritten digit recognition

Use voice to move the object

DEMO APP



"*Move the mouse to the right*"

What it does

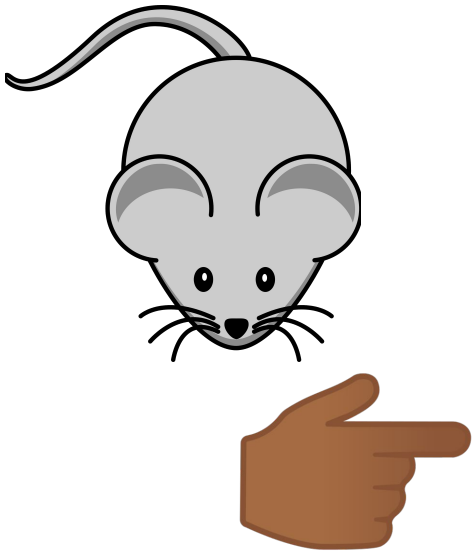
- The app is pre-trained to associate certain words with actions

How it works

- Uses a SpeechRecognition API to translate input from a microphone to text
- Only supported in Chrome & Edge browsers
- Voice data is sent in the background to Google servers to perform the conversion
- Google has pre-trained its neural network models of human speech using terabytes of voice data collected from its applications including people with impaired speech

Use gesture to move the object

DEMO APP



What it does

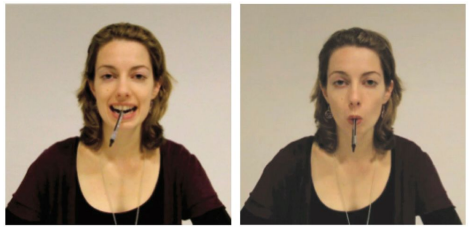
- Train the app to associate images captured by a webcam with actions

How it works

- Uses the TensorFlow.js library created by Google to bring machine learning and neural networks to the browser
- Supported by all major browsers
- Does not require network calls to a remote server

Using machine learning to enable **voice & gesture interactions in educational web applications**

Affordance: Enhance learning



- **Embodied cognition** in practice: Leverage research on how the body influences the mind to improve conceptual understanding & retention
- **Authentic assessment**: Use voice input to assess a new spoken language, or gesture input to assess sign language

Constraint: Input precision

- **Midas touch** issue: Accidental activations may relegate to low-stakes, exploratory learning tools

😸 Handwritten digit recognition has an accuracy of ~ 99%

Using machine learning to enable voice & gesture interactions in educational web applications