Actua's AI Activities Series

Activity 7

Voice Activated AI: Training Audio Recognition Models

Actua | 111 Murray Street, Ottawa, ON K1N 5M5 | www.actua.ca | 613 234 4137

Youth · STEM · Innovation
Jeunesse · STIM · Innovation

Voice Activated AI: Training Audio Recognition Models

If you're accessing this activity directly, did you know there are eight other activities in this series up on our website? If you find yourself unfamiliar with any of the AI concepts and terminology introduced in these activities, please refer to our AI Glossary. These activities also follow a space exploration narrative when done in order. It is recommended to complete the activities in order but they can also be done on their own.

You and your group mates are astronauts and scientists aboard the Actua Orbital Station. Unfortunately, your station just got bombarded by magnetic rays and your electronics have begun to shut down! The only one who can save you is the station's AI, DANN. DANN stands for Dedicated Actua Neural Network, and it's gone a little loopy. Brush up on your technical skills, learn about AI, and save yourself and your crewmates!

You've managed to reinitialize DANN's audio core using your knowledge in "What Machines See: Digging into Machine Vision", but when you asked it to open a door, it made you a milkshake instead. It looks like DANN's audio core might have sustained more damage than we thought. DANN's audio recognition model is listening to us, but it isn't understanding what it's hearing. You need to retrain it's audio processing skills so you can give them commands. Once we're sure it's fixed, we can move on to DANN's morality system in "Ethics in AI: Don't Let DANN Turn Evil"!

Activity Summary

In this activity, participants will train a machine listening program to accomplish a classification task. They will explore a pre-trained audio classifier before using Google's Teachable Machine to train their own audio-based classification model to recognize certain keywords.

Developed by Actua, 2022.

Delivery Environment	Activity Duration	Intended Audience
Classroom with computer access	60 minutes	Grades 9-12 (Ages 13-18)

Achievement Goals

Learning Goals

Learning goals are statements referring to the understanding, knowledge, skills or application participants acquire during the activity. **Following this activity, participants will:**

- **Identify and describe** one of the major applications of artificial intelligence, classification.
- **Develop and train** an artificial intelligence model for classification tasks using Google's Teachable Machine.
- **Interpret** the information reported (label and confidence) in the output of an artificial intelligence model in general terms (e.g. very confident, confident, not very confident).
- **Assess** the suitability of a trained artificial intelligence model for a defined task.

Logistics (Timing, Group Size, Materials)

Section Title	Time	Group Size	Materials
Opening Hook: Exploring a pre-trained model	10 minutes	*Entire group*	**Group Size** • Computer with microphone • Internet access: • p5 interactive sketch: Speech commands • Pen/pencil • Paper
Activity 1: Defining audio classification models	30 minutes	*Small groups or entire group*	**Group Size** • Computer with microphone • Appendix 1: RCS commands • Internet access: • Teachable Machine • p5 interactive sketch: Satellite Alignment
Reflection & Debrief	5 minutes	*Entire group*	

Safety Considerations

Safety considerations have been provided below to support safety during this activity, however they are not necessarily comprehensive. It is important that you review the activity and your delivery environment to determine any additional safety considerations that you should be implementing for the delivery of these activities.

Online Safety

Some components of this activity require the use of devices connected to the internet.

- Facilitators should review the provided videos and read/explore provided websites and materials to determine if they are suitable for their participants.
- Where applicable, facilitators should remind participants to stay on task and only use links provided within this activity.

Activity Procedure

A full reset and retraining of DANN's audio recognition model is necessary, but to accomplish that, you will need to reorient the station's main antenna array to point at the Station Recovery Satellite (SRS). Using what you learned with the visual core command model, you will need to train a simple machine listening model for handling the commands for the station's reaction control system. You can use the reaction control system to realign the antenna array to point at one of the backup satellites.

Opening Hook (time)

Just like in the "*What Machines See*" activity, you will start by exploring a pre-trained speech recognition model. This model, called SpeechCommands18w, is trained to recognize 18 words:

- The numbers "zero" to "nine"
- Four directions: "up", "down", "left", and "right"
- "Go" and "stop"
- "Yes" and "no"

This model will also indicate if it detects an "unknown word" or if it only hears "background noise". The model has been loaded into a test program for you to evaluate it: *Speech commands*

1. Open the interactive sketch. Since this is an audio recognition model, you may be asked for permission to use your computer's microphone.

2. Pick five of the words from the model's vocabulary (i.e. five of the 18 words the model has been trained to recognize).
3. Say each of the five words that you picked. Leave 1 to 2 seconds of silence between each. Write down whether or not the model successfully recognized the word that you said.
4. Repeat step 3 between 1 and 3 times.
5. Now, based on your results from steps 3 and 4, consider the following questions:
    a. Were there any words that the model had a hard time recognizing? Which ones?
    b. Did the model recognize any words incorrectly?
    c. What conditions do you think would help the model correctly recognize words?

The SpeechCommands18w model can recognize many potentially useful words, but it isn't trained to recognize the specific command words you need to use to control the station. This means that you'll need to train a model that can recognize the words that the station will respond to.

Section 1: Defining audio classification models

The process for training an audio classification model using Teachable Machine follows a similar process as training an image classification model.

Step 1. Load Teachable Machine

1. Teachable Machine can be accessed online here: *https://teachablemachine.withgoogle.com/train/audio.*
2. That link will bring you to Teachable Machine's audio classification model. You have probably already used one of the other Teachable Machine models, the image classifier, for your model for DANN's visual core.

Step 2. Define classes

1. The audio core will use the names of the classes in your model to call the corresponding commands.
2. What classes do you think you will need to define? Consult Appendix 1 for how the reaction control system (RCS) functions.
    a. *Hint: Mission Control suggests that you define one class per direction as well as the word "stop". Teachable Machine has already created the necessary "background noise" class, but it might be smart to add an "unknown word" class as in the SpeechCommands18w model.*

Step 3. Create training data

1. Training data for the audio model can be created using a microphone attached to your computer. Teachable Machine will record a continuous amount of audio (2 seconds, by default) and then automatically break it into 1 second clips for training. Each class needs a minimum of 8 different audio samples (other than the background noise class, which will need 20 samples), but more samples should mean better recognition accuracy, so do some extra if you have the time!

Step 4: Train model

Just like the image classification model that you trained, you now need to train your audio recognition model with your data so that it can recognize the required command words. To do this:

1. Click on the "Train Model" button in the box labelled "Training".
2. Wait for training to complete. After a short while, below the training button, you should see a timer counting up and a number out of 50. Your training is complete once that number reaches 50 out of 50, but this might take a few moments.
3. When training is complete, the "Preview" box should have an "Output" section which displays, in real-time, your trained model's classification of your microphone audio and your model's confidence in its classification.

Step 5: Test model

1. Now you need to check if your audio classification model's been trained well enough for you to realign the station's antenna array. This means that your model should reliably recognize the words that you chose, and detect background noise or unknown words when necessary.
2. You can adapt the evaluation questions that you used for the image classification model to test if your model is functioning well:
    a. *Does your model accurately recognize your commands when spoken by you or other group members that were part of your training data?*
    b. *Does your model accurately recognize your commands when spoken by other people who were not part of the training data?*
    c. *Does your model accurately recognize your commands when spoken in a different environment from where the training data was generated (e.g. a different part of the room, different background noises and levels)?*
3. If yes to all of the above, your model is ready for use. If not, consult the section on troubleshooting, below.
    a. Does your training data include any sounds that are not good representations of the class that they are in?
    b. Does your model work on some command classes but have difficulty recognizing specific command classes? Listening to the training data, could you hypothesize why this might be?

If the evaluation results are not satisfactory, a model can be re-trained once more training data has been added. In many cases, additional training data, or better training data, will solve issues of model reliability and accuracy. A model that can reliably pass at least one of the evaluation questions may still be loaded into the visual core at the discretion of the user. Reliable performance, however, cannot be guaranteed.

Step 6. Applying your trained model

P5 Interactive Model: Satellite Alignment

Paste the URL for your Teachable Machine model where the website prompts you to. This will load your AI model into DANN's audio core.

1. Use your model to reorient the station's antenna array and align.

Reflection & Debrief

Once you've uploaded your trained audio model into the interactive sketch and used it, discuss the following questions:

1. Did your model function well for its purpose?
2. Were there any words that your model had trouble understanding?
3. What do you think the pros and cons are of voice interfaces?
    a. Where do we see them applied?
    b. Where might they be helpful?
4. If you've completed the previous activity for training an image classification model: Does your audio model work better than your image classification model? Support your position with specific observations and examples.

Delivery Adaptations

How might you adapt the time, space, materials, group sizes, or instructions to make this activity more approachable or more challenging? **Modifications** are ways to make the activity more accessible, **extensions** are ways to make the activity last longer or more challenging.

Modifications

- For the opening hook model exploration, to save time, instead of doing it in small groups, it can be done as a whole class with students taking turns saying words from the vocabulary.
- To make the activity easier for younger participants, Teachable Machine supports pre-recording a dataset and uploading it to Google Drive. This allows participants to build on an existing data set for less complex experience and better results.

Extensions

- "Model and data provenance" are information about the model's creation and training. Review the model and data provenance section for the SpeechCommands18w model, here: *ML5.js*
    - Why is this information important?
    - How could this information help us build better models?
- This activity can be extended by creating new classes and audio commands. Imagine that you are giving commands to an AI in charge of a space station. What commands would you need to give it? Create and train new classes with those commands, making sure that you produce the same amount of data as you did for the other classes. Test it out. When the model has more classes, is it better or worse at identifying commands? How could it be improved?

References & Gratitude

This activity was made possible thanks to Teachable Machine, which can be found at *https://teachablemachine.withgoogle.com/train/audio.*

Several activity pieces were built using p5.js, an online javascript library. It can be found at *https://p5js.org/.*

Terms of Use

Prior to using this activity or parts thereof, you agree and understand that:

- It is your responsibility to review all aspects of this activity and ensure safety measures are in place for the protection of all involved parties.
- Any safety precautions contained in the “Safety Considerations” section of this write-up are not intended as a complete list or to replace your own safety review process.
- Actua shall not be responsible or liable for any damage that may occur due to your use of this content.
- This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. For more information, please see https://creativecommons.org/licenses/by-nc-sa/4.0/.
- You may adapt the content for your program (remix, transform, and build upon the material), providing appropriate credit to Actua and indicating if changes were made. No sharing of content with third parties without written permission from Actua.

About Actua

Actua is Canada's leading science, technology, engineering and mathematics (STEM) youth outreach network, representing a growing network of over 40 universities and colleges across the country. Each year 350,000 young Canadians in over 500 communities nationwide are inspired through hands-on educational workshops, camps and community outreach initiatives. Actua focuses on the engagement of underrepresented youth through specialized programs for Indigenous youth, girls and young women, at-risk youth and youth living in Northern and remote communities. For more information, please visit us online at www.actua.ca and on social media: Twitter, Facebook, Instagram and YouTube!

Appendices

Appendix A: Career & Mentor Connections

- Audio Engineer
- Data Analyst
- Machine Learning Analyst
- Computer Programmer

Appendix B: Background Information

RCS COMMANDS

The commands that the RCS is expecting are:

- PORT and STBD (pronounced “starboard”):
    - PORT activates the port-side thruster which rotates the state clockwise (or right).
    - STBD activates the starboard-side thruster which rotates the station counterclockwise (or left).
- POS (short for “positive”) and NEG (short for “negative”) pitch thrusters:
    - POS activates the positive pitch thruster which tilts the station and the antenna array upward.
    - NEG activates the negative pitch thruster which tilts the station and the antenna array downward.
- STOP:
    - STOP instructs the station to take control of the thrusters to bring any motion to a halt.

The command name is in ALL CAPS and each command executes for one second, except for the STOP command which takes as long as necessary to bring the station motion to a halt. Once rotating, the station will not stop naturally. An opposing thruster will need to be fired to stop the station’s motion.