



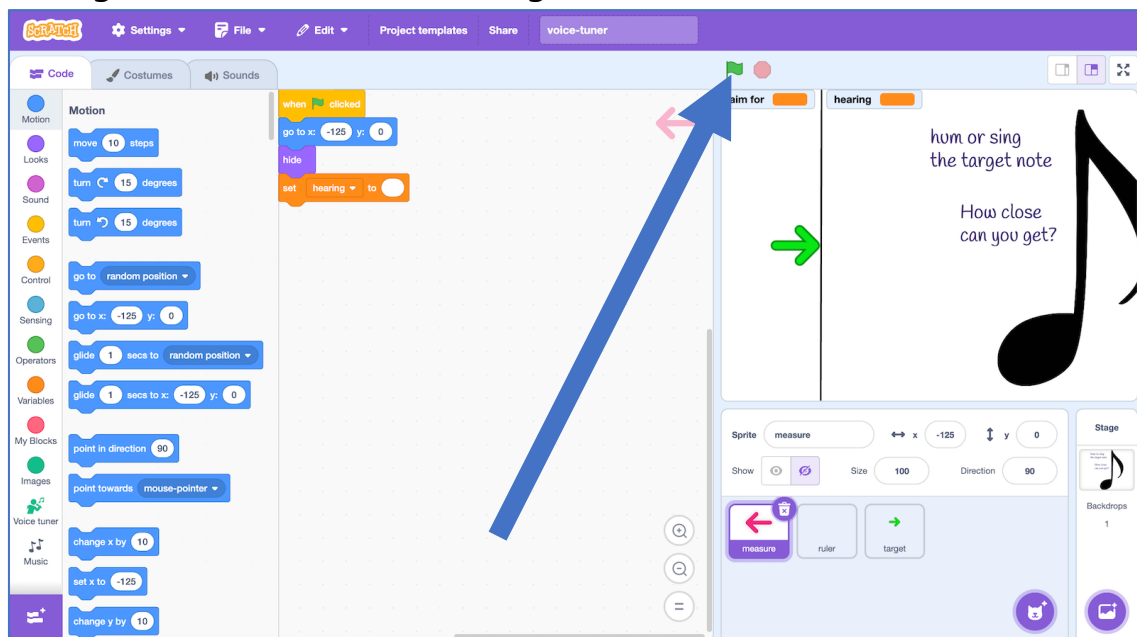
Page 1 of 5

Last updated: 6 December 2024

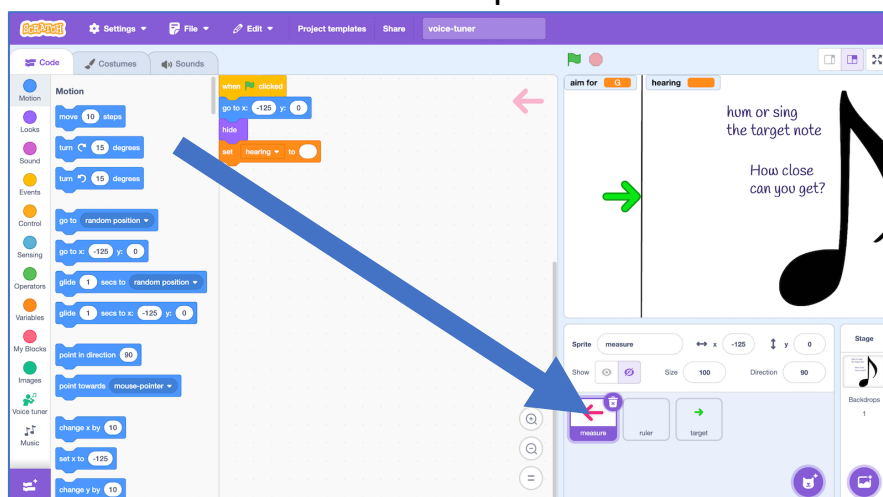


This project worksheet is licensed under a Creative Commons Attribution Non-Commercial Share-Alike License  
<http://creativecommons.org/licenses/by-nc-sa/4.0/>

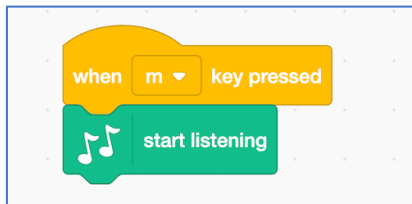
1. Go to <https://machinelearningforkids.co.uk/scratch>
2. Click on “Project templates”
3. Click on “Voice Tuner” and wait for the project to download
4. Click on the **Green Flag**  
*Scratch will choose a random musical note and play it to you.  
Your goal will be to hum or sing that note.*



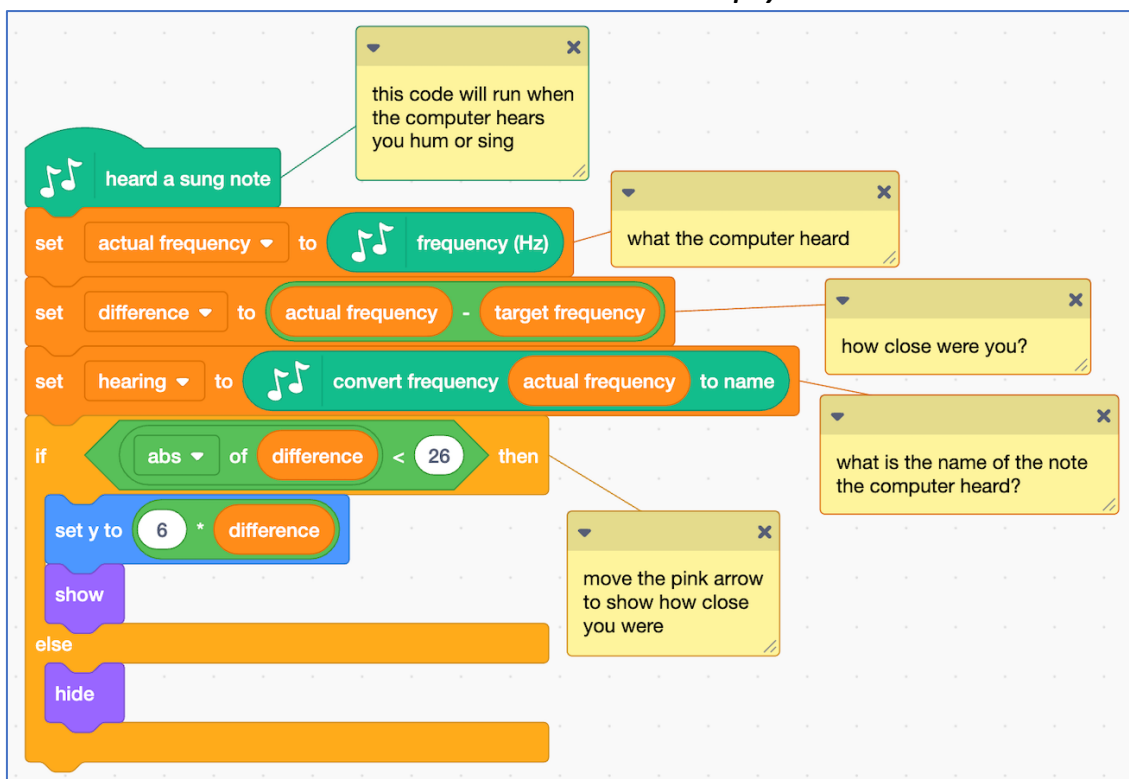
5. *If you want to see how the random note is being chosen and played, look at the code in the **target** sprite*
6. Click on the **measure** sprite



7. Create this code, so Scratch will start listening to your humming and singing when you press the **M** key on your keyboard  
*You can pick a different key if you like. I chose “M” for “Music”.*



8. Create this code  
*You don't need to add the comments.*  
*The comments are in the screenshot to help you understand the code.*

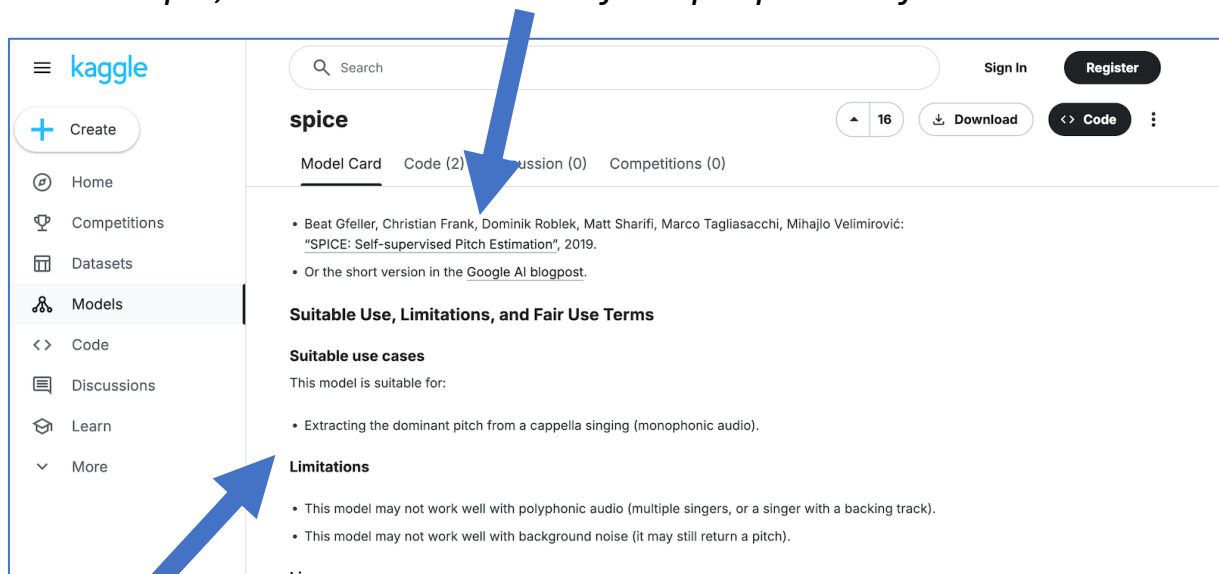


9. Give it a try!  
*Click the Green Flag to choose a new note and hear what it sounds like.*  
*Press the **M** key to start Scratch listening. Try to hum or sing the note.*

10. **How many** recordings of people's voices do you think were collected to train this model?  
**How** do you think the recordings were collected?  
*Try to guess before you move on to the next step.*  
*Write down your guesses to help you remember.*

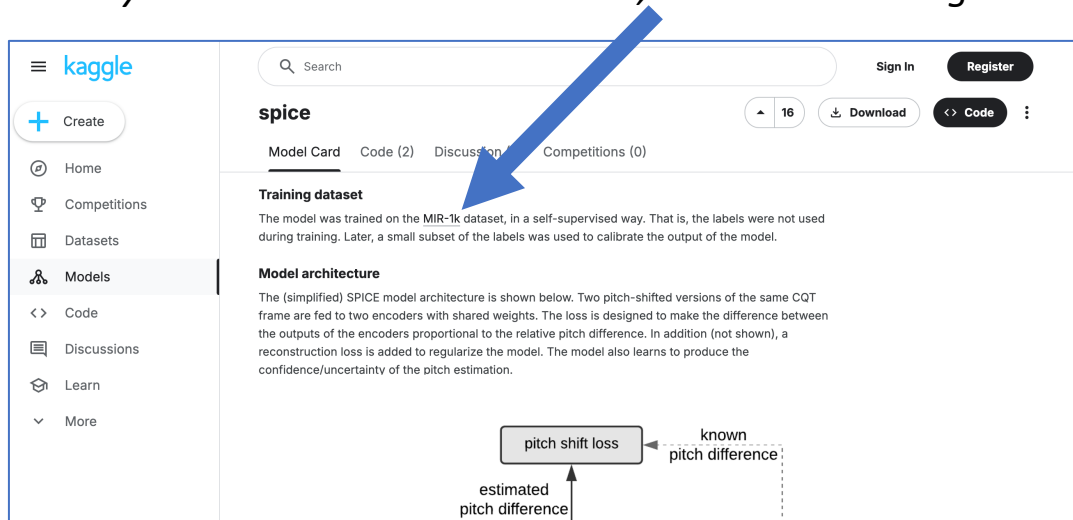
**11.** In a new web browser window, go to <https://ibm.biz/spiceml>  
*When we use a machine learning system, it is helpful if the creators of the system publish the “Model card” – where they tell you how it was created. The model you have been using is **SPICE**. This web page has the model card for this machine learning model.*

**12.** Have a look at the model card  
*Model cards are complicated, so don’t worry that you won’t understand all of it. But there are some interesting things to find here.  
For example, here are the names of the people who first created it.*

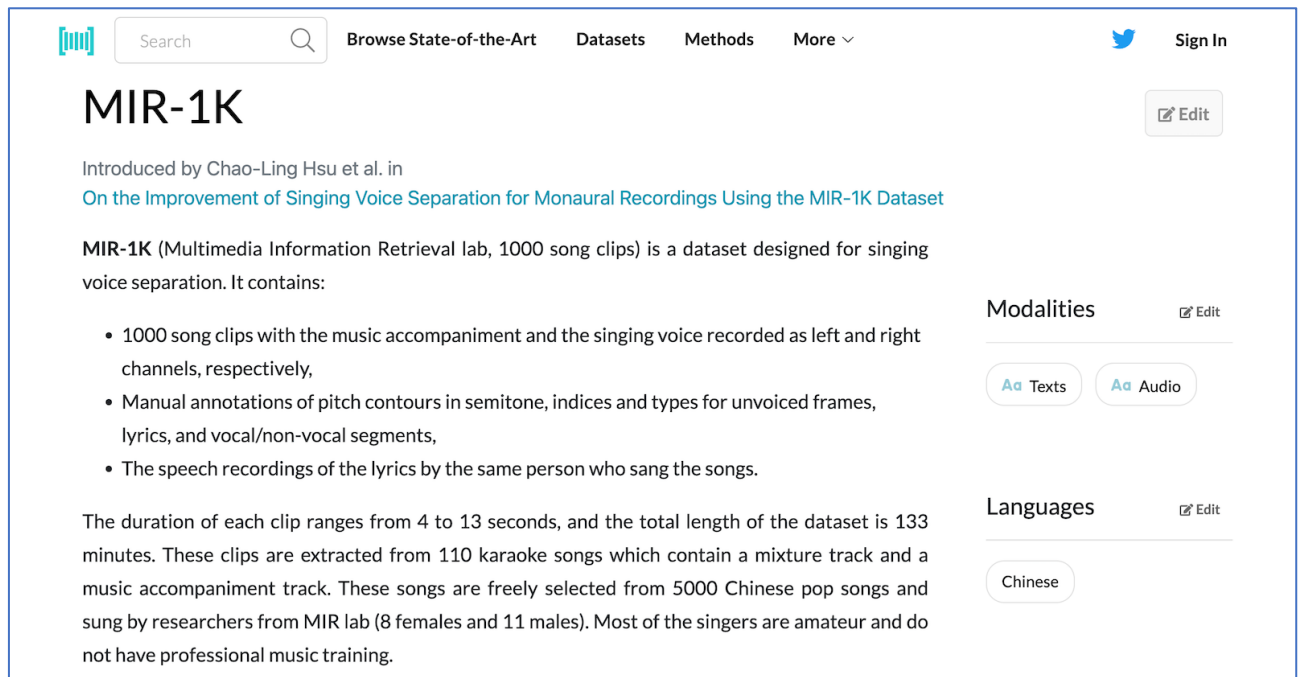


*This is telling you that the model has been trained to recognise the note from the sound of a single voice, without background music.*

*This says how the model was trained, and what training data was used.*



You can find information about the MIR-1k training set at <https://paperswithcode.com/dataset/mir-1k>



The screenshot shows the MIR-1K dataset page on the Papers with Code website. The page has a header with a search bar, navigation links (Browse State-of-the-Art, Datasets, Methods, More), and a Sign In button. The main content area features the title 'MIR-1K' with an 'Edit' button. Below the title, it states the dataset was introduced by Chao-Ling Hsu et al. in the paper 'On the Improvement of Singing Voice Separation for Monaural Recordings Using the MIR-1K Dataset'. A description of the dataset follows: 'MIR-1K (Multimedia Information Retrieval lab, 1000 song clips) is a dataset designed for singing voice separation. It contains:'. A bulleted list details the dataset's contents: 1000 song clips with music accompaniment and singing voice, manual annotations of pitch contours and lyrics, and speech recordings of lyrics. A paragraph then describes the clip duration (4-13 seconds), total length (133 minutes), and source (110 karaoke songs from 5000 Chinese pop songs). On the right, there are sections for 'Modalities' (Texts, Audio) and 'Languages' (Chinese), each with an 'Edit' button.

**MIR-1K** Edit

Introduced by Chao-Ling Hsu et al. in  
[On the Improvement of Singing Voice Separation for Monaural Recordings Using the MIR-1K Dataset](#)

**MIR-1K** (Multimedia Information Retrieval lab, 1000 song clips) is a dataset designed for singing voice separation. It contains:

- 1000 song clips with the music accompaniment and the singing voice recorded as left and right channels, respectively,
- Manual annotations of pitch contours in semitone, indices and types for unvoiced frames, lyrics, and vocal/non-vocal segments,
- The speech recordings of the lyrics by the same person who sang the songs.

The duration of each clip ranges from 4 to 13 seconds, and the total length of the dataset is 133 minutes. These clips are extracted from 110 karaoke songs which contain a mixture track and a music accompaniment track. These songs are freely selected from 5000 Chinese pop songs and sung by researchers from MIR lab (8 females and 11 males). Most of the singers are amateur and do not have professional music training.

**Modalities** Edit

Texts Audio

**Languages** Edit

Chinese

*This tells you that MIR-1k contains **one thousand recordings** of computing researchers singing Chinese songs in a karaoke!*

*Compare that with your guess. Were you close?*

## What have you done?

You've been using a machine learning model to estimate the pitch of a human voice.

You've learned what a "model card" is. Not all artificial intelligence systems publish a model card, but it is a good practice that encourages transparency.

You have seen one way that the training sets needed to create real-world machine learning models are collected.

**What else could you make with this machine learning model?**

## Ideas and Extensions

Now that you've finished, why not give one of these ideas a try?

Or come up with one of your own?