**Surrey Audio-Visual Expressed Emotion (SAVEE)**

<https://www.kaggle.com/datasets/barelydedicated/savee-database>

Speakers:

'DC', 'JE', 'JK' and 'KL' are four male speakers recorded for the SAVEE database

Audio data :

Audio files consist of audio WAV files sampled at 44.1 kHz

There are 15 sentences for each of the 7 emotion categories. Total dataset consists of 480 files.

The initial letter(s) of the file name represents the emotion class, and the following digits represent the sentence number.

The letters 'a', 'd', 'f', 'h', 'n', 'sa' and 'su' represent 'anger', 'disgust', 'fear', 'happiness', 'neutral', 'sadness' and 'surprise' emotion classes respectively.

E.g., 'd03.wav' is the 3rd disgust sentence.

**Toronto emotional speech set (TESS)**

<https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess>

There are a set of 200 target words were spoken in the carrier phrase "Say the word **\_**' by two actresses (aged 26 and 64 years) and recordings were made of the set portraying each of seven emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral).

Each emotion is segregated into different directories, hence total of 14 directories (7 emotions, 2 actresses)

There are 2800 data points (audio files) in total.

The dataset is organised such that each of the two female actor and their emotions are contain within its own folder. And within that, all 200 target words audio file can be found. The format of the audio file is a WAV format

**Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)**

<https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio>

**Files**

This portion of the RAVDESS contains 1440 files: 60 trials per actor x 24 actors = 1440. The RAVDESS contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent. Speech emotions includes calm, happy, sad, angry, fearful, surprise, and disgust expressions. Each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression.

**File naming convention**

Each of the 1440 files has a unique filename. The filename consists of a 7-part numerical identifier (e.g., 03-01-06-01-02-01-12.wav). These identifiers define the stimulus characteristics:

*Filename identifiers*

* Modality (01 = full-AV, 02 = video-only, 03 = audio-only).
* Vocal channel (01 = speech, 02 = song).
* Emotion (01 = neutral, 02 = calm, 03 = happy, 04 = sad, 05 = angry, 06 = fearful, 07 = disgust, 08 = surprised).
* Emotional intensity (01 = normal, 02 = strong). NOTE: There is no strong intensity for the 'neutral' emotion.
* Statement (01 = "Kids are talking by the door", 02 = "Dogs are sitting by the door").
* Repetition (01 = 1st repetition, 02 = 2nd repetition).
* Actor (01 to 24. Odd numbered actors are male, even numbered actors are female).

*Filename example: 03-01-06-01-02-01-12.wav*

1. Audio-only (03)
2. Speech (01)
3. Fearful (06)
4. Normal intensity (01)
5. Statement "dogs" (02)
6. 1st Repetition (01)
7. 12th Actor (12)  
   Female, as the actor ID number is even.