

DATASET

Dewansh Kr Singh

dewansh@iitk.ac.in

(+91)7906806147

1. OVERVIEW

The speech samples have been taken from three different sources

1. 2800 files from **TESS** - These 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). There are 2800 data points (audio files) in total. The dataset is organized 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

License Information:

Toronto Emotional Speech Set License Attribution NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

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

2. 1440 Speech samples + 1012 Song samples from **RAVDESS** - The RAVDESS contains 24 professional actors (12 females, 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.

License Information:

"The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)" by Livingstone & Russo is licensed under CC BY-NA-SC 4.0.

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

<https://www.kaggle.com/uwrkaggler/ravdess-emotional-song-audio>

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
3. 480 files from **SAVEE** - The SAVEE database was recorded from four native English male speakers (identified as DC, JE, JK, KL), postgraduate students and researchers at the University of Surrey aged from 27 to 31 years.

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

Content

This results in a total of 120 utterances per speaker, for example

Common: She had your dark suit in greasy wash water all year.

Anger: Who authorized the unlimited expense account?

Disgust: Please take this dirty table cloth to the cleaners for me.

Fear: Call an ambulance for medical assistance.

Happiness: Those musicians harmonize marvelously.

Sadness: The prospect of cutting back spending is an unpleasant one for any governor.

Surprise: The carpet cleaners shampooed our oriental rug.

Neutral: The best way to learn is to solve extra problems.

(*This dataset is skewed since there is no Calm class in both the SAVEE and TESS database and the folder in actor 18 for song sample is empty)

2. CONVERSION –

Both the databases from the TESS and SAVEE datasets have been renamed as per the RAVDESS naming convention for easy extraction of features from them. One can simply use the data already converted to proper naming convention or can create it from scratch himself/herself with the steps below:

- Two python code files called **savee.py** and **tess.py** have been created and presented in the same folder as this document.
- Simply download the TESS and SAVEE database from the links above.
- Copy both the python code files to the same folder where you downloaded the datasets and simply execute savee.py and tess.py
- Each code will create two additional folders TESS_RAVDESS and SAVEE_RAVDESS with the respective renamed files that now can be used with RAVDESS files easily.

File name change rules (in accordance to RAVDESS dataset)

TESS	SAVEE
<ul style="list-style-type: none"> • Modality (03) • Vocal channel (01 = speech) • Emotion ({1to8}/{2}) • Emotional intensity (random) • Statement (random) • Repetition (random) • Actor (OAF: 25, YAF: 26) 	<ul style="list-style-type: none"> • Modality (03) • Vocal channel (01 = speech) • Emotion ({1to8}/{2}) • Emotional intensity (random) • Statement (random) • Repetition (random) • Actor (DC:27, JE:29, JK:31, KL=33)

The RAVDESS dataset has been manually split into RAVDESS_split1 and RAVDESS_split2 so as to make two sets with no overlapping speaker.