

## Emotion Diarization with WavLM Large on 5 popular emotional datasets.

This repository provides all the necessary tools to perform speech emotion diarization with a fine-tuned wavlm (large) model using SpeechBrain.

The model is trained on concatenated audios and tested on [ZaionEmotionDataset](#). The metric is Emotion Diarization Error Rate (EDER). For more details please check the [paper link](#).

For a better experience, we encourage you to learn more about [SpeechBrain](#). The model performance on ZED (test set) is:

Release	EDER(%)
05-07-23	29.7 (Avg: 30.2)

### Pipeline description

This system is composed of a wavlm encoder a downstream frame-wise classifier. The task aims to predict the correct emotion components and their boundaries within a speech recording. For now, the model was trained with audios that contain only 1 non-neutral emotion event.

The system is trained with recordings sampled at 16kHz (single channel). The code will automatically normalize your audio (i.e., resampling + mono channel selection) when calling `diarize_file` if needed.

### Install SpeechBrain

First of all, please install the **development** version of SpeechBrain with the following command:

```
git clone https://github.com/speechbrain/speechbrain.git
cd speechbrain
pip install -r requirements.txt
pip install --editable .
```

Please notice that we encourage you to read our tutorials and learn more about [SpeechBrain](#).

### Perform Speech Emotion Diarization

```
from speechbrain.inference.diarization import Speech_Emotion_Diarization
classifier = Speech_Emotion_Diarization.from_hparams(
    source="speechbrain/emotion-diarization-wavlm-large"
)
diary = classifier.diarize_file("speechbrain/emotion-diarization-wavlm-large/example.wav")
print(diary)

# {
#   'speechbrain/emotion-diarization-wavlm-large/example.wav':
#     [
#       {'start': 0.0, 'end': 1.94, 'emotion': 'n'}, # n -> neutral
#       {'start': 1.94, 'end': 4.48, 'emotion': 'h'} # h -> happy
#     ]
# }

diary = classifier.diarize_file("speechbrain/emotion-diarization-wavlm-large/example_sad.wav")
print(diary)

# {
#   'speechbrain/emotion-diarization-wavlm-large/example_sad.wav':
#     [
#       {'start': 0.0, 'end': 3.54, 'emotion': 's'}, # s -> sad
#       {'start': 3.54, 'end': 5.26, 'emotion': 'n'} # n -> neutral
#     ]
# }
```

The output will contain a dictionary of emotion components and their boundaries.

### Inference on GPU

To perform inference on the GPU, add `xun_opts={"device": "cuda"}` when calling the `from_hparams` method.

### Training

The model was trained with SpeechBrain (aa018540). To train it from scratch follow these steps:

- Clone SpeechBrain:

```
git clone https://github.com/speechbrain/speechbrain/
```

- Install it:

```
cd speechbrain
pip install -r requirements.txt
pip install -e .
```

- Run Training:

```
cd recipes/ZaionEmotionDataset/emotion_diarization
python train.py hparams/train.yaml --zed_folder /path/to/ZED --emovdb_folder /path/to/emovdb
```

You can find our training results (models, logs, etc) [here](#).

### Limitations

The SpeechBrain team does not provide any warranty on the performance achieved by this model when used on other datasets.

### About Speech Emotion Diarization/Zaion Emotion Dataset

```
@article{wang2023speech,
  title={Speech Emotion Diarization: Which Emotion Appears When?},
  author={Wang, Yingzhi and Ravanelli, Mirco and Nfissi, Alaa and Yacoubi, Alya},
  journal={arXiv preprint arXiv:2306.12991},
  year={2023}
}
```

### Citing SpeechBrain

Please, cite SpeechBrain if you use it for your research or business.

```
@misc{speechbrain,
  title={SpeechBrain: A General-Purpose Speech Toolkit},
  author={Mirco Ravanelli and Titouan Parcollet and Peter Plantinga and Aku Rouhe},
  year={2021},
  eprint={2106.04624},
  archivePrefix={arXiv},
  primaryClass={eess.AS},
  note={arXiv:2106.04624}
}
```

### About SpeechBrain

- Website: <https://speechbrain.github.io/>
- Code: <https://github.com/speechbrain/speechbrain/>
- HuggingFace: <https://huggingface.co/speechbrain/>

Edit model card

Downloads last month  
**477**



### Inference Examples

Audio Classification

Inference API (serverless) has been turned off for this model.