report

August 31, 2021

1 VAD

1.1 Abstract

1.2 Introduction

1.3 Method

1.3.1 Dataset

All audio files were associated with a label file of the same name. 1. Volume 1. 1914 files 1. 957 .wav audio 2. 957 .json abels

1.3.2 Technical setup

1.3.3 Model implementation

System description

Neural network architecture

Input

GRU layer

Dense layer

Batch inference

1.4 Experiments

1.4.1 Experimental setup

1.5 Results

1.5.1 Several speakers

Listening to a sample of the audio files revealed that a variety of speakers

- Humans
  - men
  - women

- Synthetic
  - men
  - women

I also characterised speeches by their variety of amplitudes and pace - Normal vs fast pace
- Loud vs, low volume

1.5.2 Speech signal description

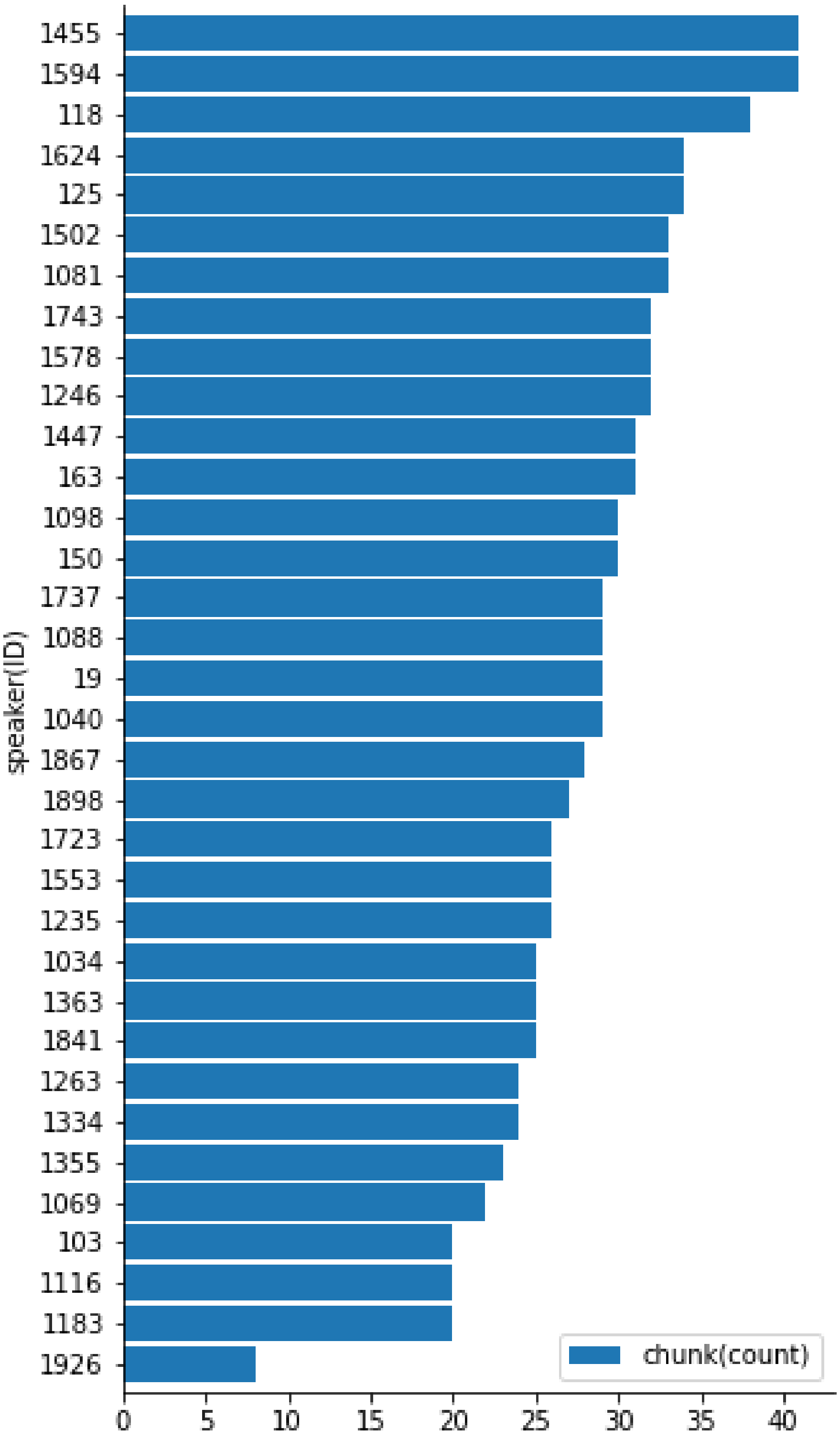
We show below the best typical example of an audio signal (top panel). and its associated speech labels “1” for speech and “0” for no speech (bottom panel).

All audio signals were 32 bits float single channel time series. We run a few sanity checks:

```
- the 957 label files were correctly mapped with the 957 audio files
```

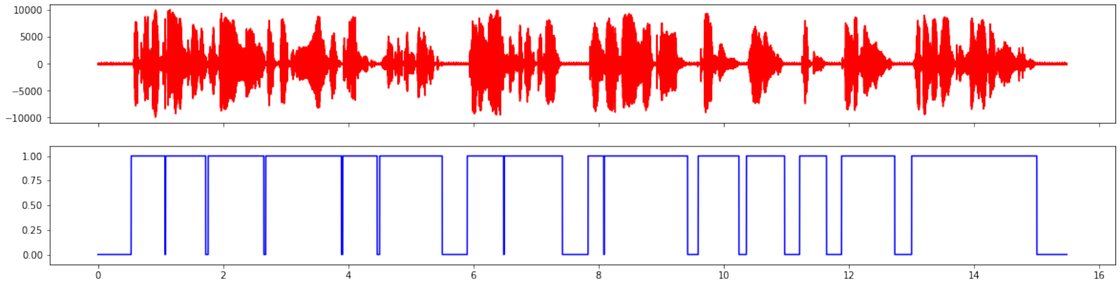
Speaker information:

```
- Number: 34 speakers
- Speakers'ID: ['103' '1034' '1040' '1069' '1081' '1088' '1098' '1116' '118'
'1183'
 '1235' '1246' '125' '1263' '1334' '1355' '1363' '1447' '1455' '150'
 '1502' '1553' '1578' '1594' '1624' '163' '1723' '1737' '1743' '1841'
 '1867' '1898' '19' '1926']
```

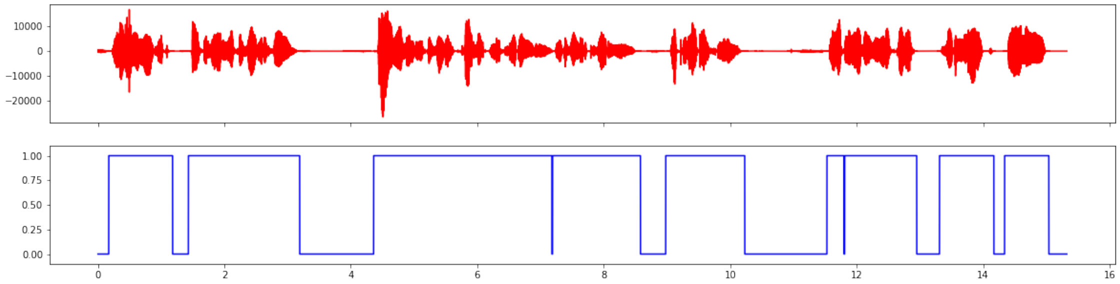


We show below a few interesting example chunks for 7 different speakers (numbered panels). - All visualized audio were very well labelled (see supplementary).

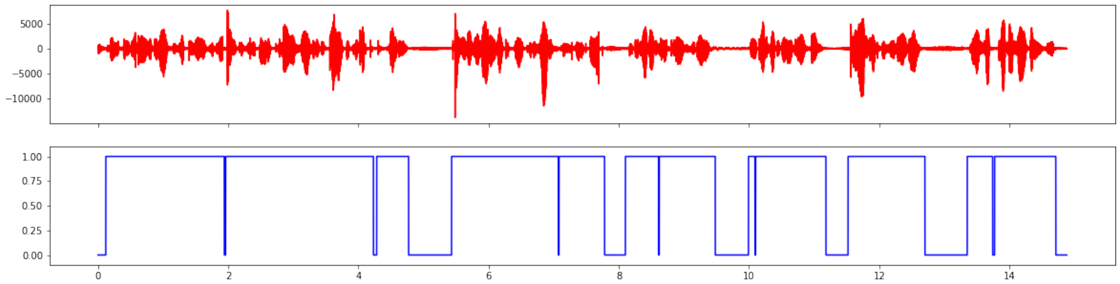
```
SPEAKER 19 - PANEL 0
  data/01_raw/vad_data/19-198-0003.wav
  data/01_raw/vad_data/19-198-0003.json
```



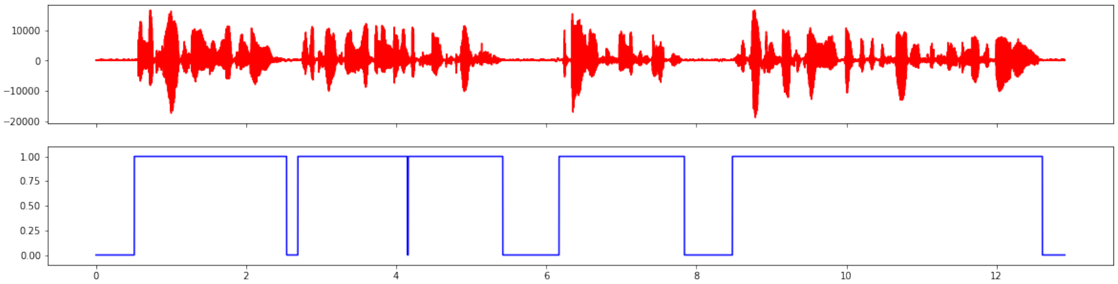
```
SPEAKER 1553 - PANEL 1
  data/01_raw/vad_data/1553-140048-0009.wav
  data/01_raw/vad_data/1553-140048-0009.json
```



```
SPEAKER 103 - PANEL 2
  data/01_raw/vad_data/103-1241-0027.wav
  data/01_raw/vad_data/103-1241-0027.json
```



```
SPEAKER 1034 - PANEL 3
  data/01_raw/vad_data/1034-121119-0047.wav
  data/01_raw/vad_data/1034-121119-0047.json
```



```
2.06 sec
```

We validated that all audio files were associated with a .json label file.

```
- audio file sample size: 957
- label file sample size: 957
```

The entire sample could be quickly loaded:

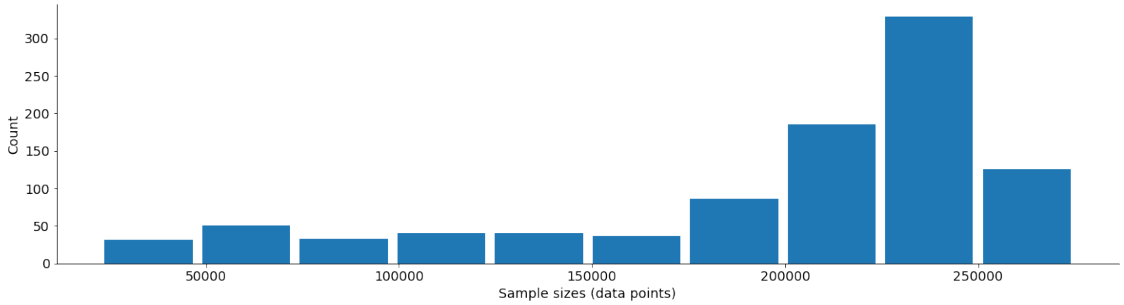
```
- loading duration: 2.36 sec
```

Sample size and sampling rate:

```
Sample rate information:
- 1 sample rate(s)
- rate: 16000 Hz
```

We kept the signal at 16Khz which is enough to cover the frequency range of human speech according to the literature (Human voice b/w `85hz to 8khz` [REF], hearing b/w `20 hz to 20kh`[REF]).

```
Sample size information:
- 711 sample size(s)
- max: 275280 samples ( [17.205] secs)
- min: 22560 samples ( [1.41] secs)
- median: 222080.0 samples ( [13.88] secs)
```



Signal amplitudes: the true decibel amplitude of the audio will depend on each speaker's microphone characteristics, the speaker's distance to its microphone, the speaker's volume configuration. Having no acces to these information we did not derive the true decibel amplitude (dB) from the raw audio signal amplitude or compared absolute amplitudes between speakers. Rather we compared the signals' signal-to-noise ratio (SNR).

1.5.3 Speech signals are nearly pure

N_rms is the root-mean square level of the noise without speech.

Average audio duration

1.6 Conclusion

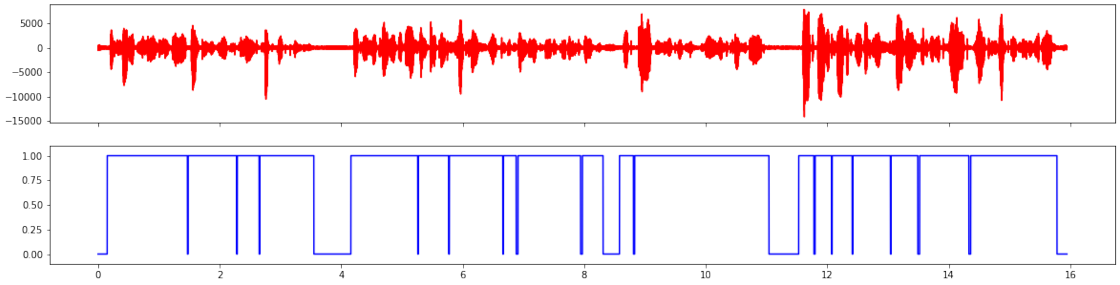
1.7 Discussion

1.8 References

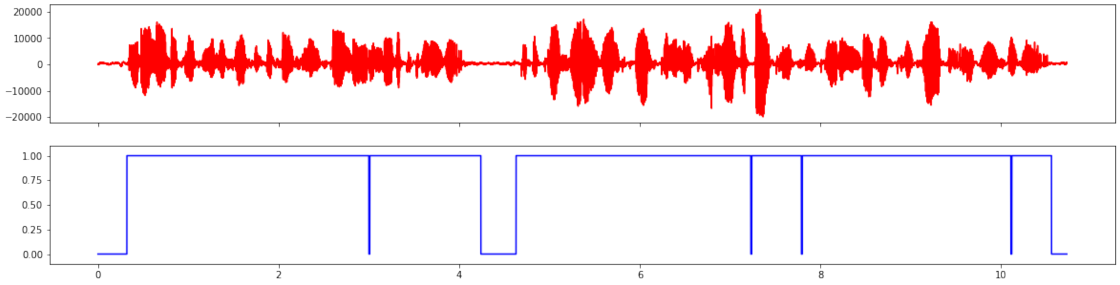
1.9 Supplementary results

1.10 Each speaker first audio signal

```
SPEAKER 103 - PANEL 0
  data/01_raw/vad_data/103-1240-0001.wav
  data/01_raw/vad_data/103-1240-0001.json
```

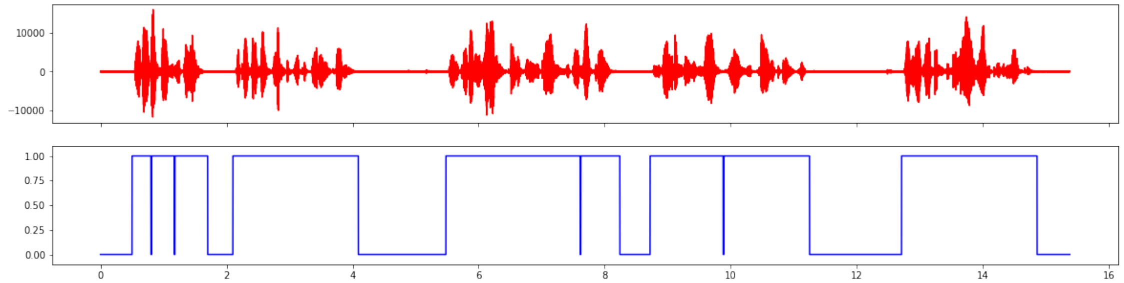


```
SPEAKER 1034 - PANEL 1
  data/01_raw/vad_data/1034-121119-0005.wav
  data/01_raw/vad_data/1034-121119-0005.json
```

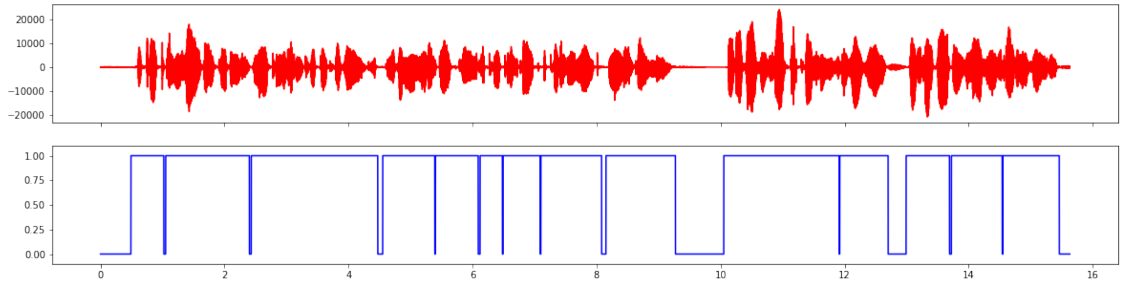


```
SPEAKER 1040 - PANEL 2
  data/01_raw/vad_data/1040-133433-0001.wav
```

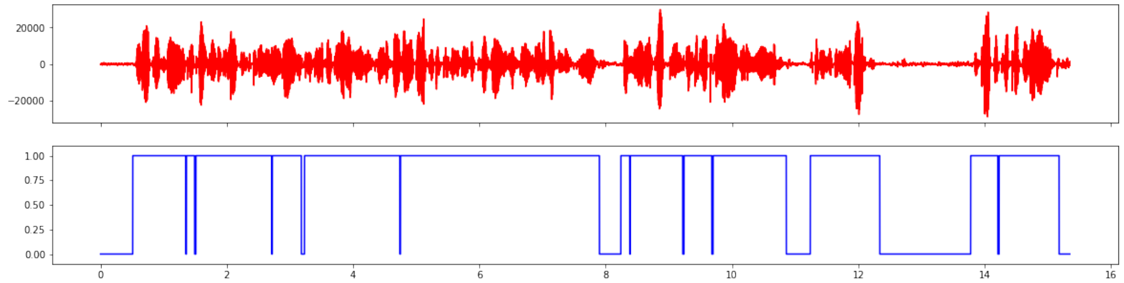
data/01_raw/vad_data/1040-133433-0001.json



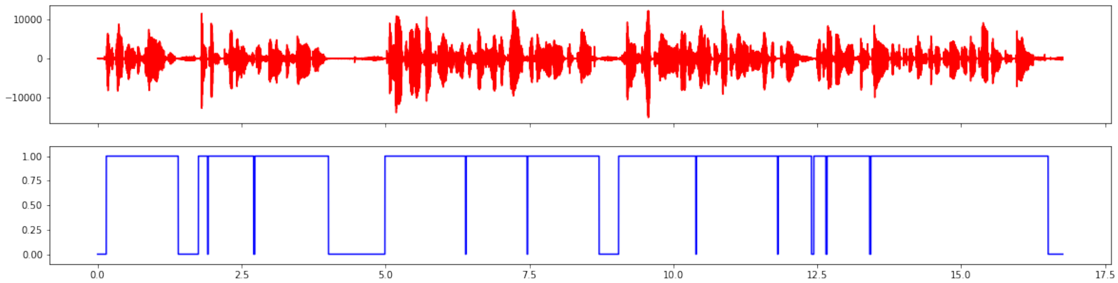
SPEAKER 1069 - PANEL 3
data/01_raw/vad_data/1069-133699-0000.wav
data/01_raw/vad_data/1069-133699-0000.json



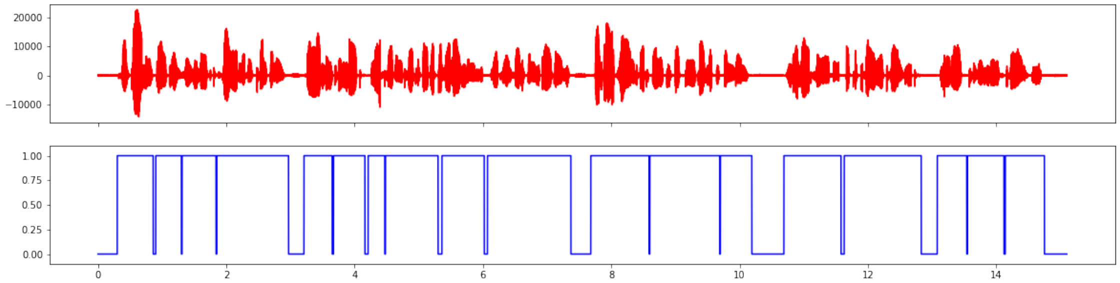
SPEAKER 1081 - PANEL 4
data/01_raw/vad_data/1081-125237-0007.wav
data/01_raw/vad_data/1081-125237-0007.json



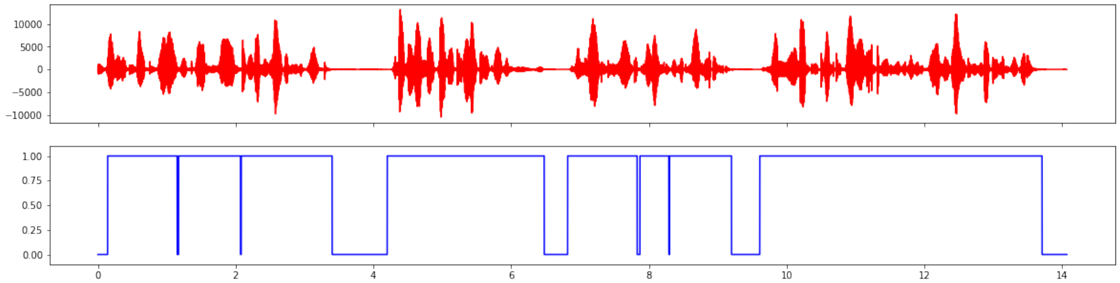
SPEAKER 1088 - PANEL 5
data/01_raw/vad_data/1088-129236-0003.wav
data/01_raw/vad_data/1088-129236-0003.json



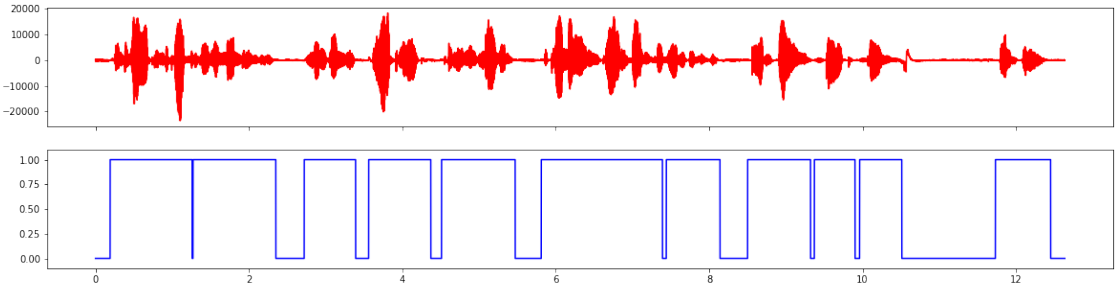
SPEAKER 1098 - PANEL 6
  data/01_raw/vad_data/1098-133695-0001.wav
  data/01_raw/vad_data/1098-133695-0001.json



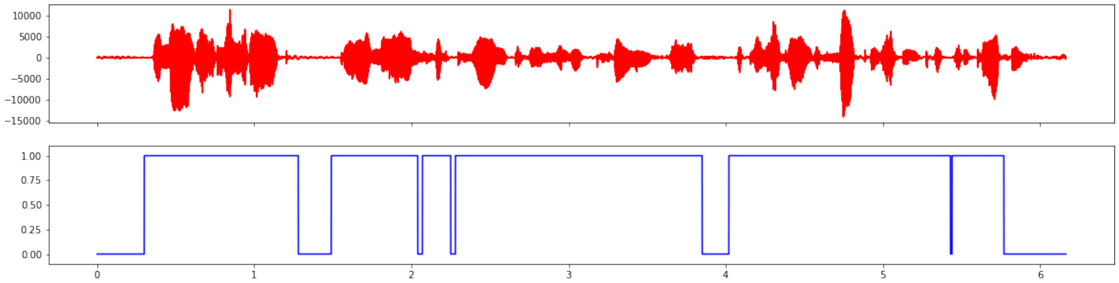
SPEAKER 1116 - PANEL 7
  data/01_raw/vad_data/1116-132847-0003.wav
  data/01_raw/vad_data/1116-132847-0003.json



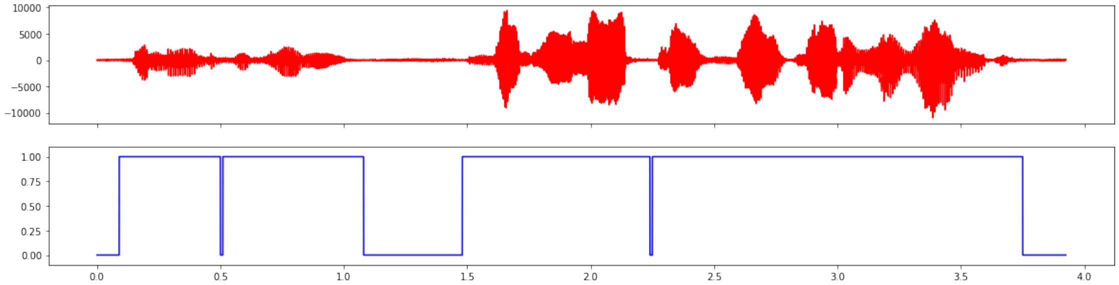
SPEAKER 118 - PANEL 8
  data/01_raw/vad_data/118-121721-0005.wav
  data/01_raw/vad_data/118-121721-0005.json



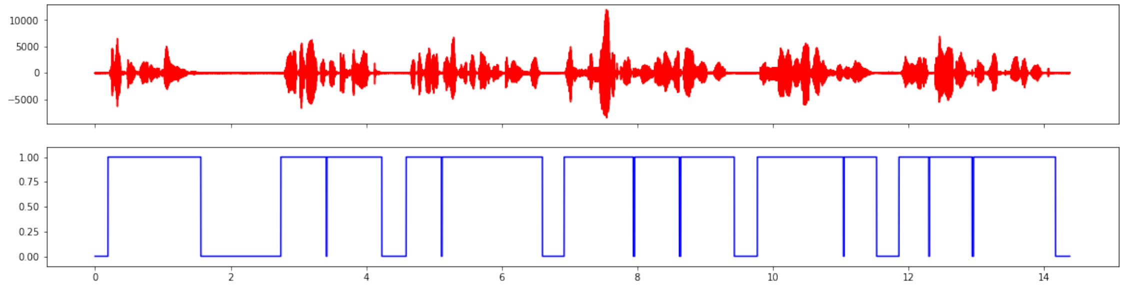
SPEAKER 1183 - PANEL 9
  data/01_raw/vad_data/1183-124566-0000.wav
  data/01_raw/vad_data/1183-124566-0000.json



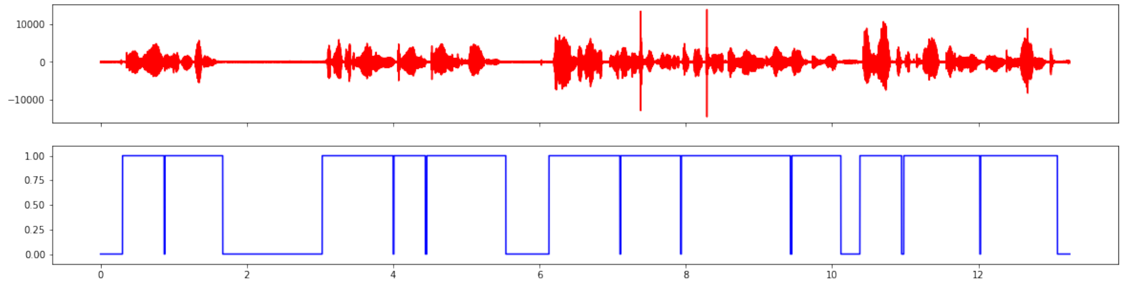
SPEAKER 1235 - PANEL 10
  data/01_raw/vad_data/1235-135883-0007.wav
  data/01_raw/vad_data/1235-135883-0007.json



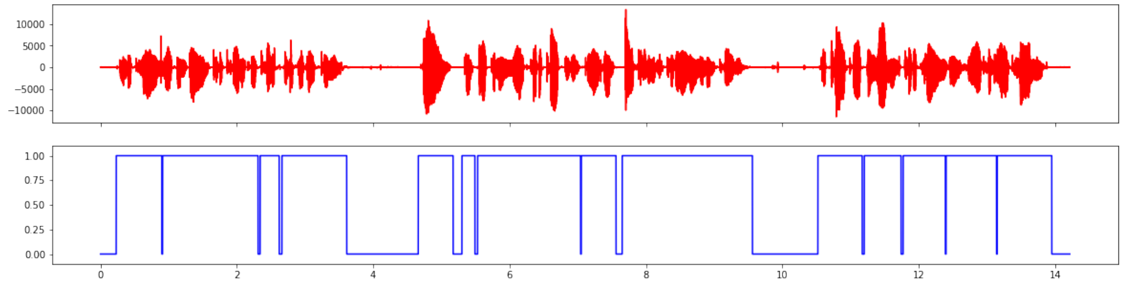
SPEAKER 1246 - PANEL 11
  data/01_raw/vad_data/1246-124548-0000.wav
  data/01_raw/vad_data/1246-124548-0000.json



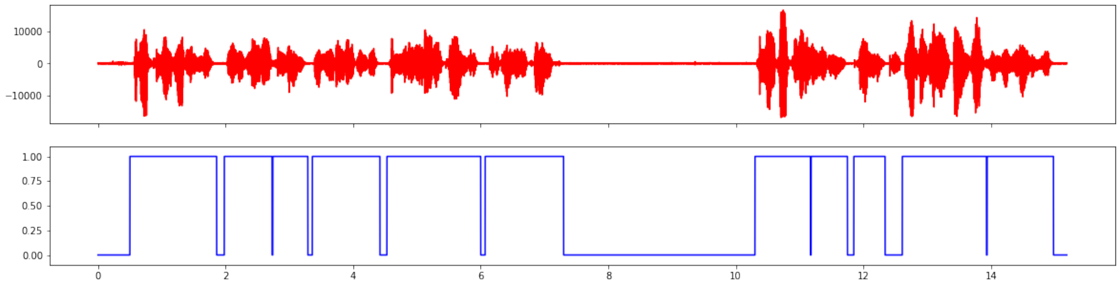
```
SPEAKER 125 - PANEL 12
  data/01_raw/vad_data/125-121124-0000.wav
  data/01_raw/vad_data/125-121124-0000.json
```



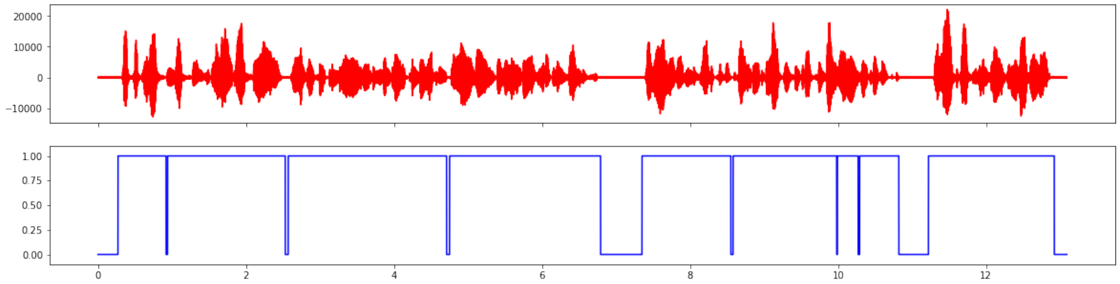
```
SPEAKER 1263 - PANEL 13
  data/01_raw/vad_data/1263-138246-0000.wav
  data/01_raw/vad_data/1263-138246-0000.json
```



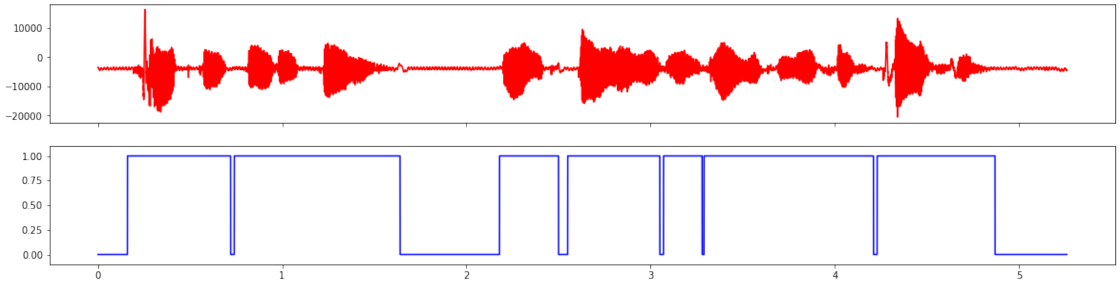
```
SPEAKER 1334 - PANEL 14
  data/01_raw/vad_data/1334-135589-0011.wav
  data/01_raw/vad_data/1334-135589-0011.json
```



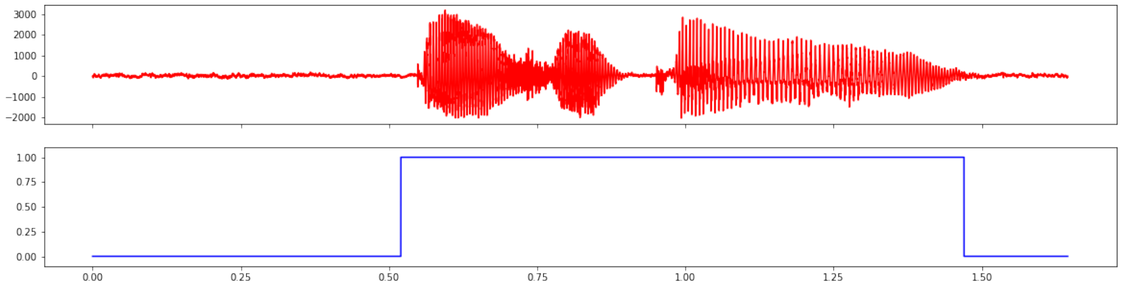
```
SPEAKER 1355 - PANEL 15
  data/01_raw/vad_data/1355-39947-0014.wav
  data/01_raw/vad_data/1355-39947-0014.json
```



```
SPEAKER 1363 - PANEL 16
  data/01_raw/vad_data/1363-135842-0000.wav
  data/01_raw/vad_data/1363-135842-0000.json
```

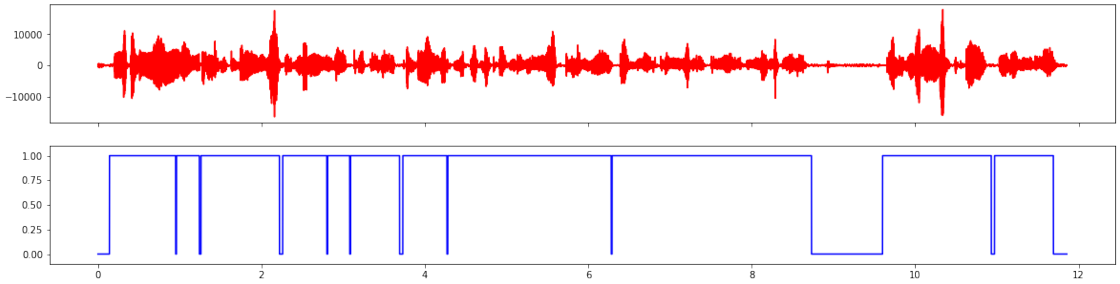


```
SPEAKER 1447 - PANEL 17
  data/01_raw/vad_data/1447-130550-0000.wav
  data/01_raw/vad_data/1447-130550-0000.json
```



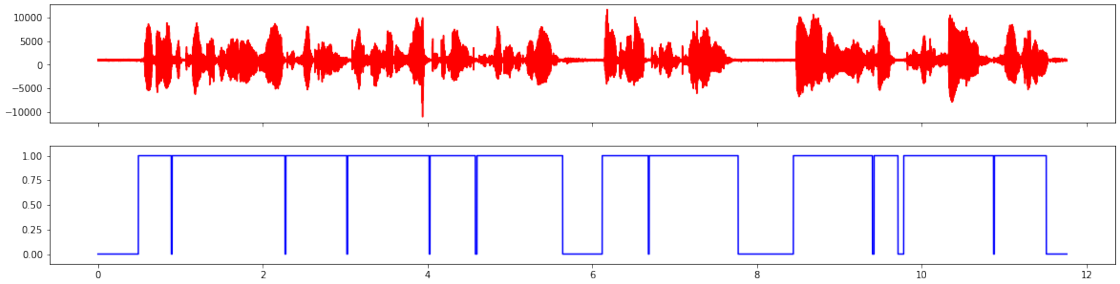
SPEAKER 1455 - PANEL 18

```
  data/01_raw/vad_data/1455-134435-0007.wav
  data/01_raw/vad_data/1455-134435-0007.json
```



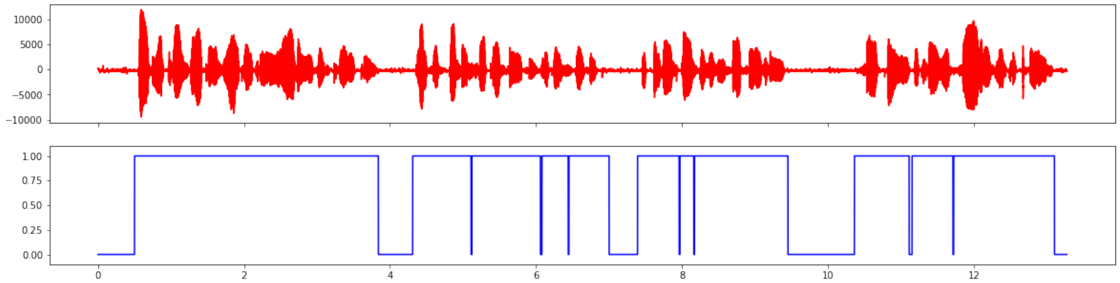
SPEAKER 150 - PANEL 19

```
  data/01_raw/vad_data/150-126107-0001.wav
  data/01_raw/vad_data/150-126107-0001.json
```

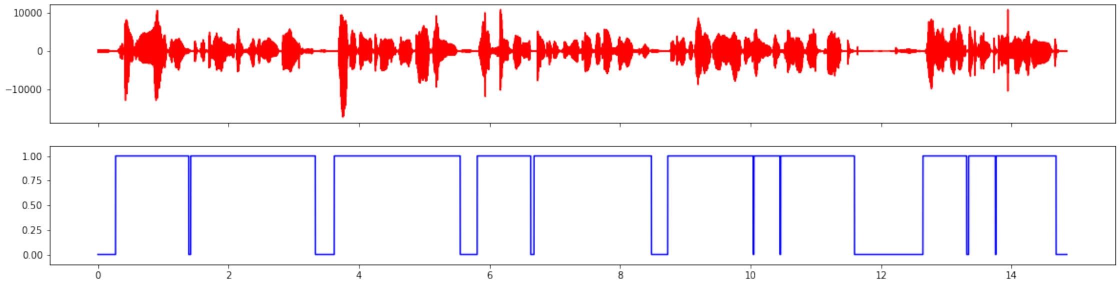


SPEAKER 1502 - PANEL 20

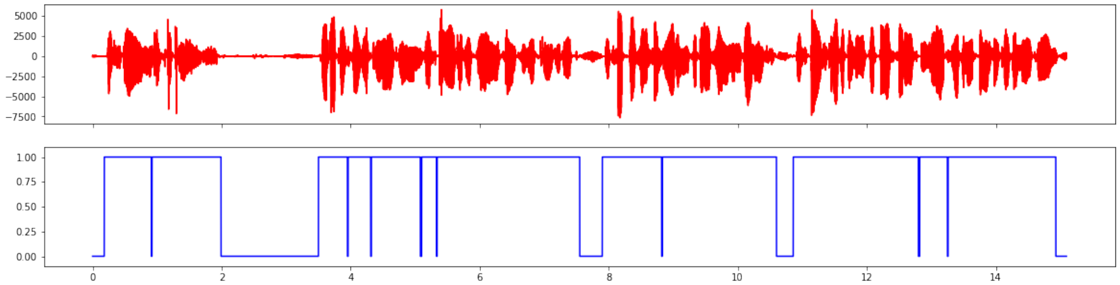
```
  data/01_raw/vad_data/1502-122615-0007.wav
  data/01_raw/vad_data/1502-122615-0007.json
```



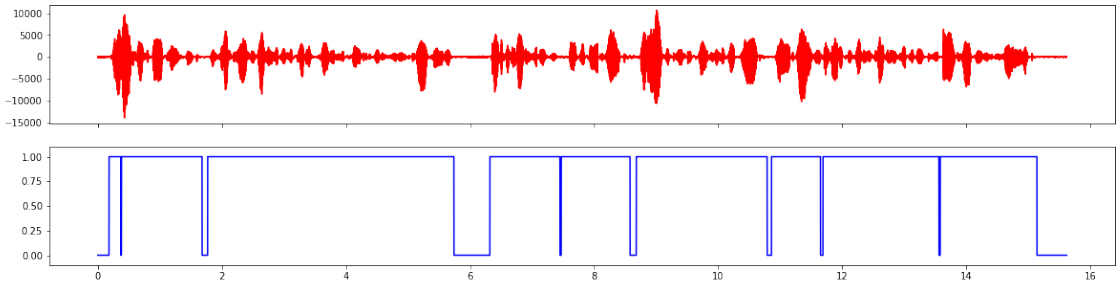
SPEAKER 1553 - PANEL 21
  data/01_raw/vad_data/1553-140047-0002.wav
  data/01_raw/vad_data/1553-140047-0002.json



SPEAKER 1578 - PANEL 22
  data/01_raw/vad_data/1578-140045-0000.wav
  data/01_raw/vad_data/1578-140045-0000.json

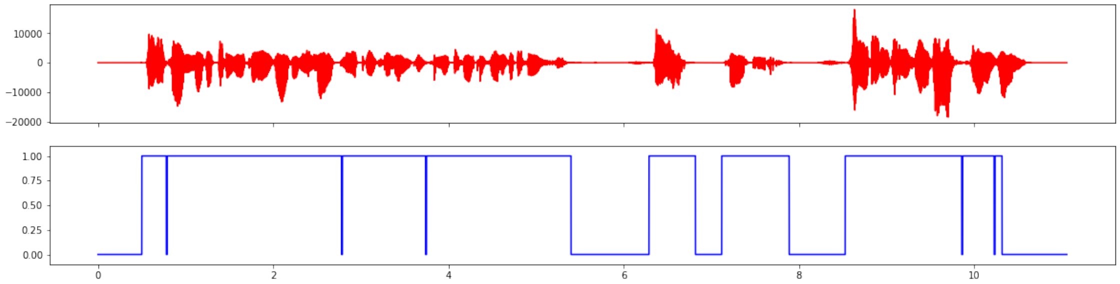


SPEAKER 1594 - PANEL 23
  data/01_raw/vad_data/1594-135914-0004.wav
  data/01_raw/vad_data/1594-135914-0004.json



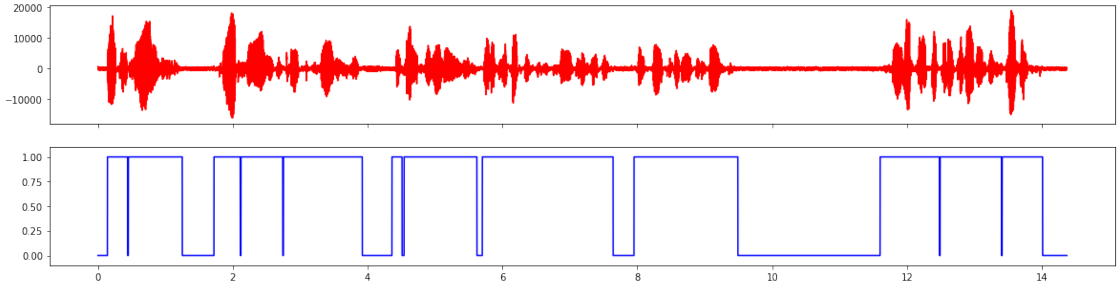
SPEAKER 1624 - PANEL 24

```
  data/01_raw/vad_data/1624-142933-0003.wav
  data/01_raw/vad_data/1624-142933-0003.json
```



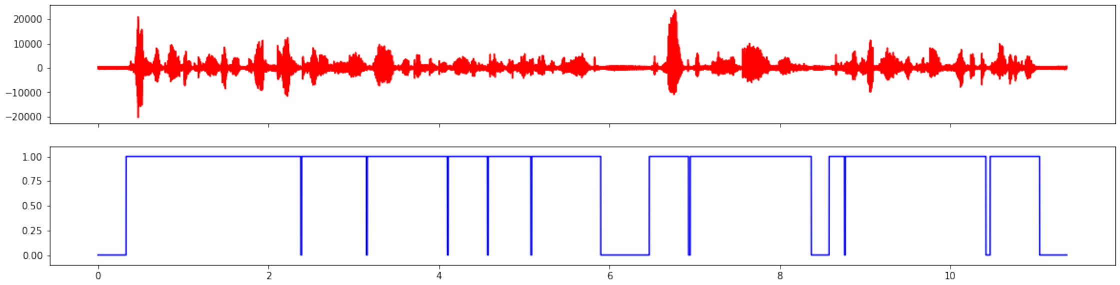
SPEAKER 163 - PANEL 25

```
  data/01_raw/vad_data/163-121908-0006.wav
  data/01_raw/vad_data/163-121908-0006.json
```

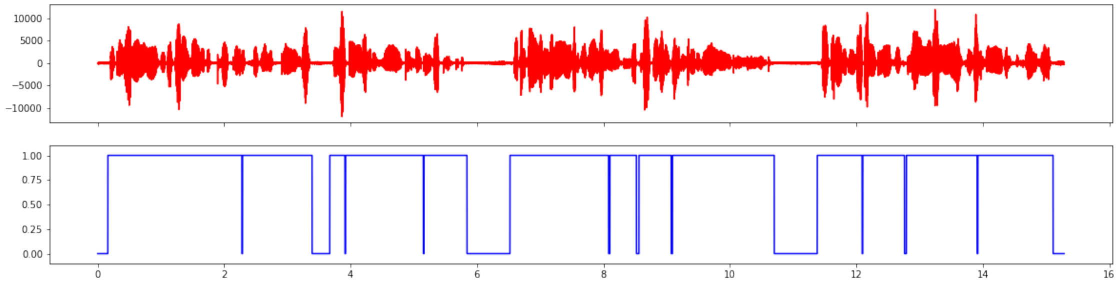


SPEAKER 1723 - PANEL 26

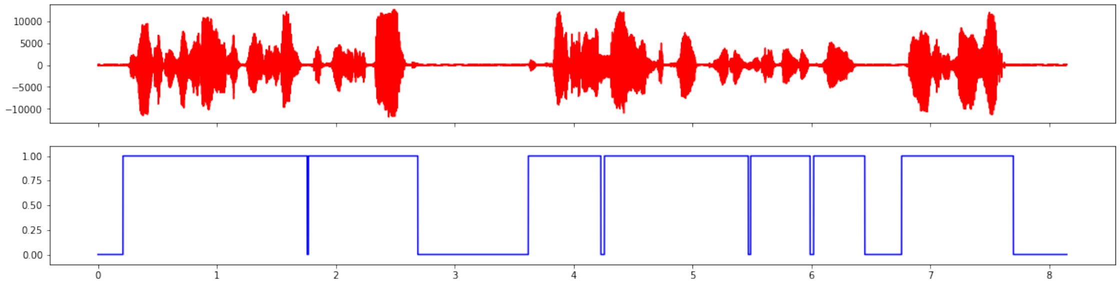
```
  data/01_raw/vad_data/1723-141149-0005.wav
  data/01_raw/vad_data/1723-141149-0005.json
```



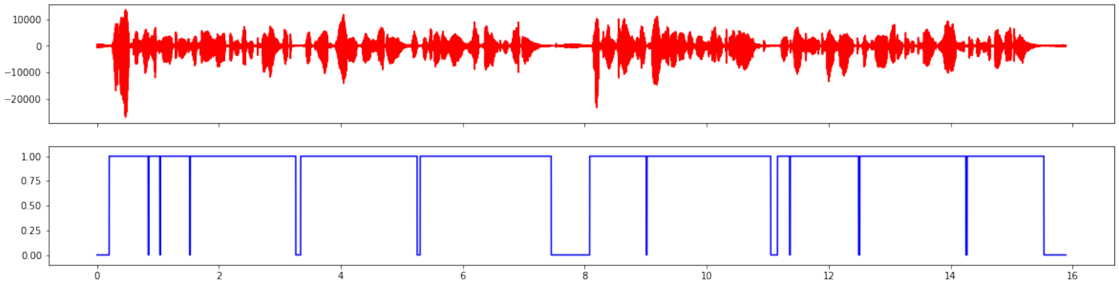
SPEAKER 1737 - PANEL 27
  data/01_raw/vad_data/1737-142396-0000.wav
  data/01_raw/vad_data/1737-142396-0000.json



SPEAKER 1743 - PANEL 28
  data/01_raw/vad_data/1743-142912-0002.wav
  data/01_raw/vad_data/1743-142912-0002.json

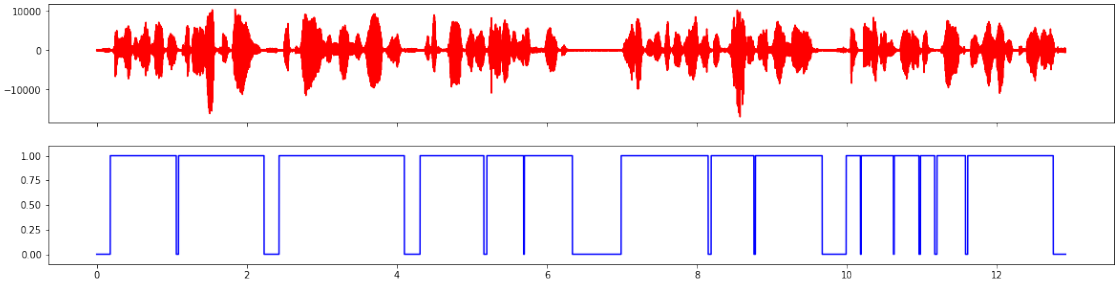


SPEAKER 1841 - PANEL 29
  data/01_raw/vad_data/1841-150351-0013.wav
  data/01_raw/vad_data/1841-150351-0013.json



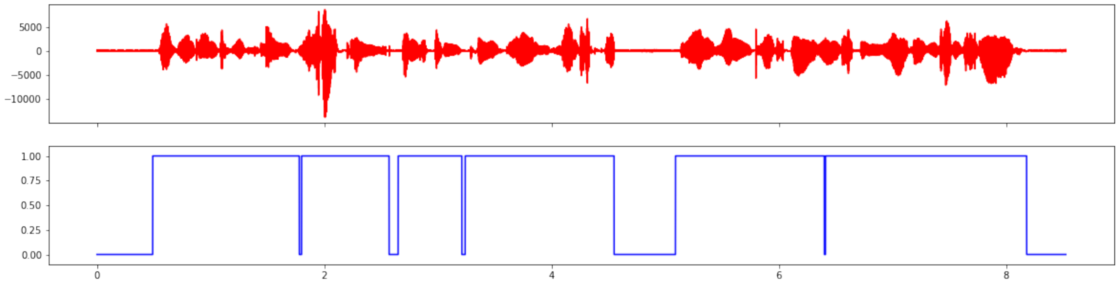
SPEAKER 1867 - PANEL 30

```
  data/01_raw/vad_data/1867-148436-0001.wav
  data/01_raw/vad_data/1867-148436-0001.json
```



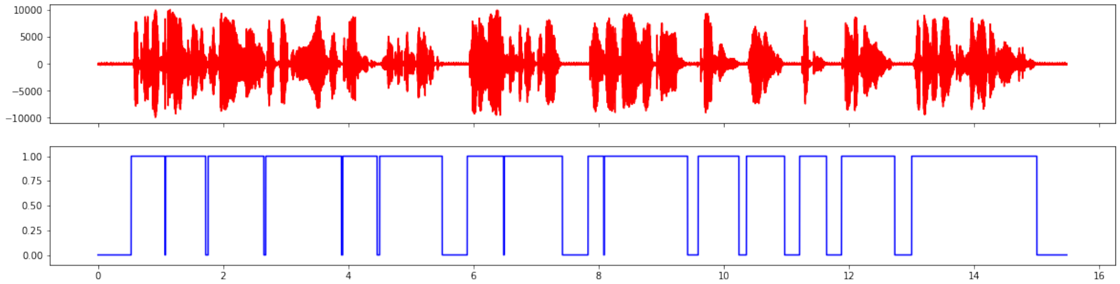
SPEAKER 1898 - PANEL 31

```
  data/01_raw/vad_data/1898-145702-0007.wav
  data/01_raw/vad_data/1898-145702-0007.json
```

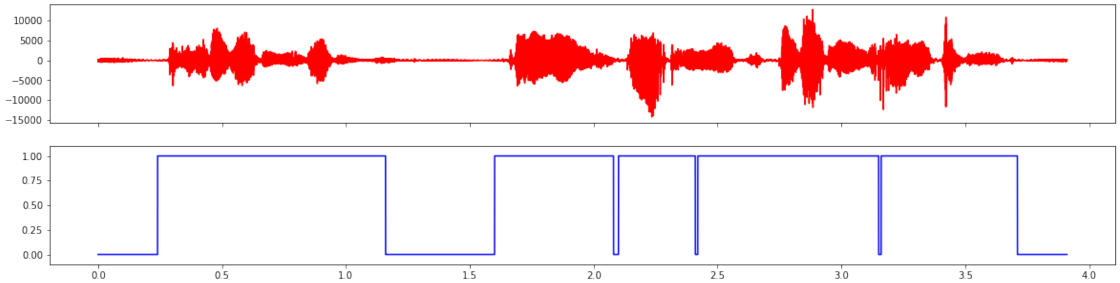


SPEAKER 19 - PANEL 32

```
  data/01_raw/vad_data/19-198-0003.wav
  data/01_raw/vad_data/19-198-0003.json
```



```
SPEAKER 1926 - PANEL 33
  data/01_raw/vad_data/1926-143879-0002.wav
  data/01_raw/vad_data/1926-143879-0002.json
```



```
18.41 sec
```