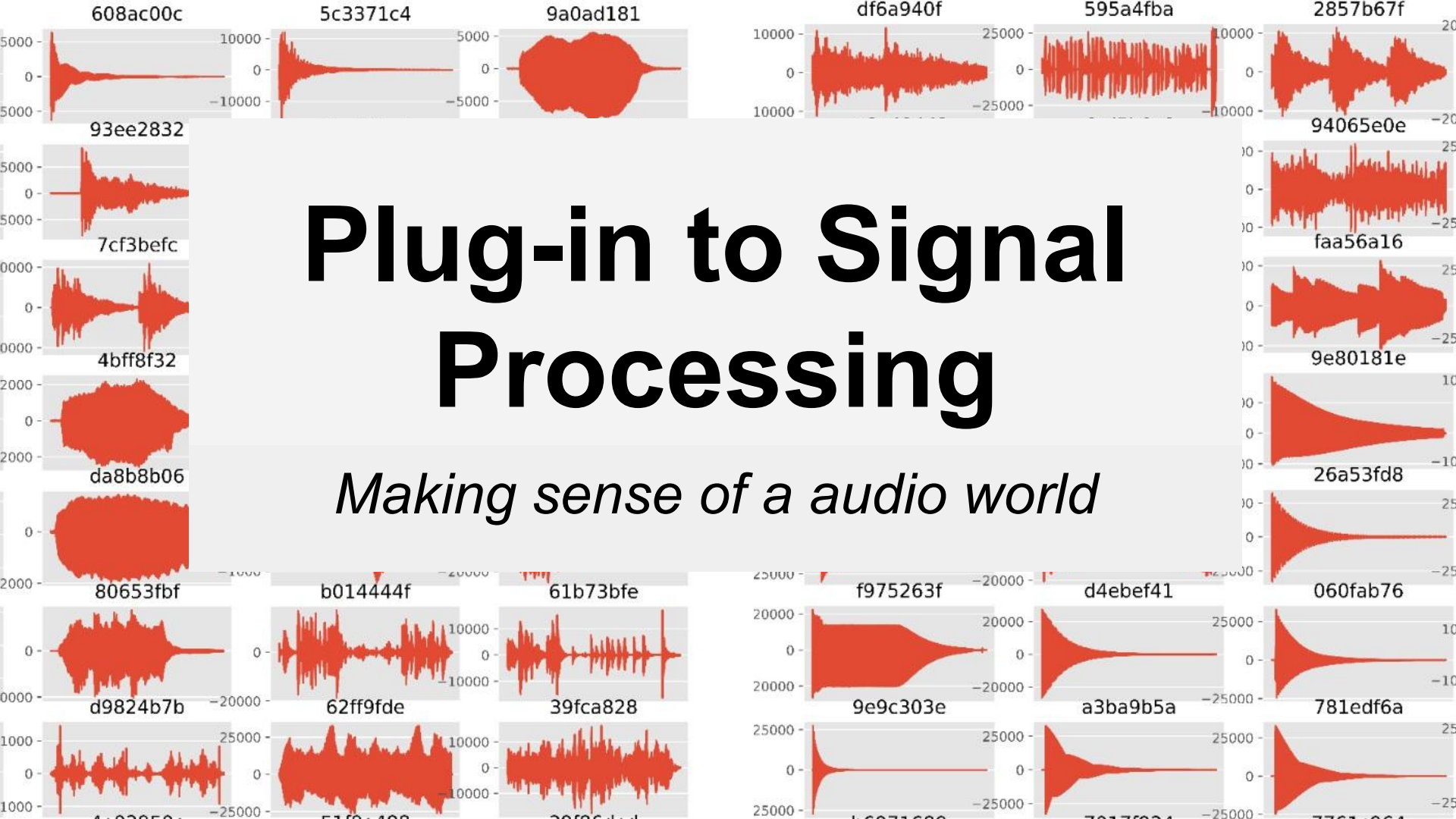
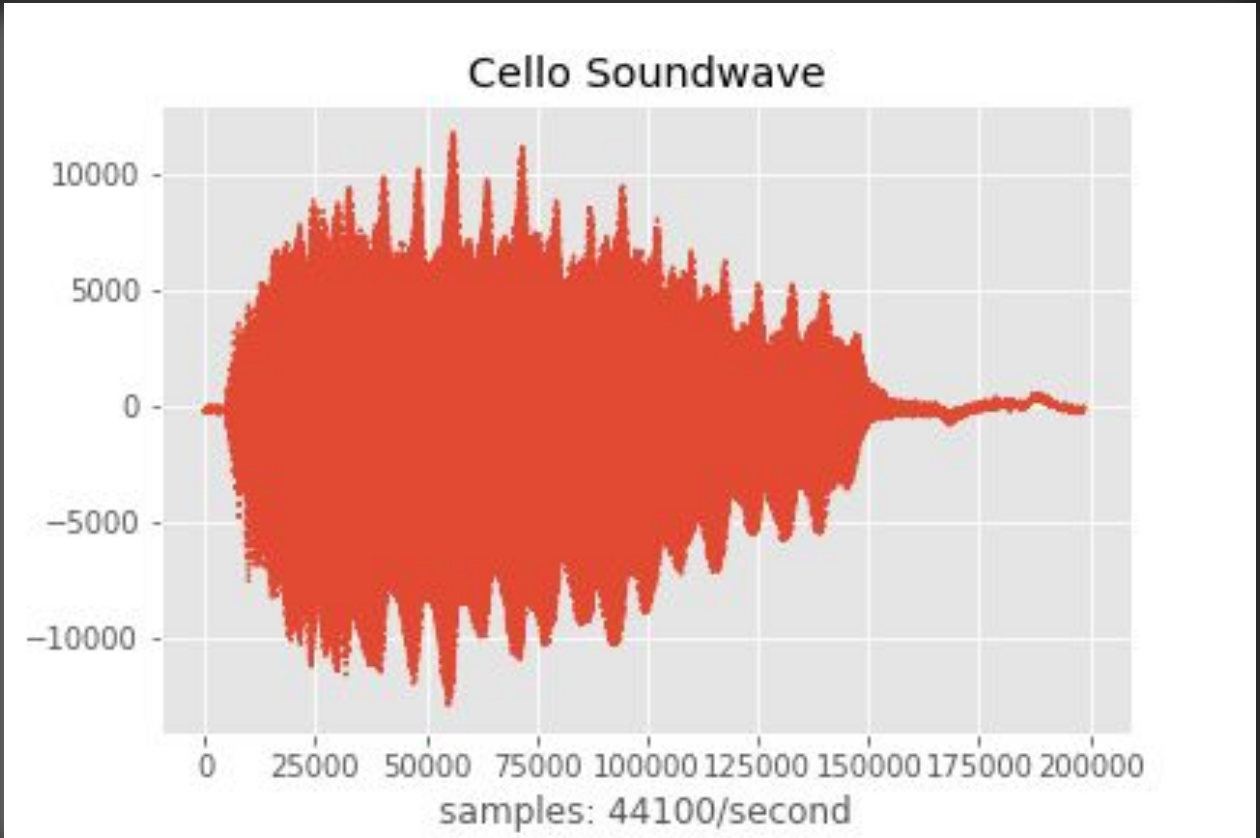
Goo
l



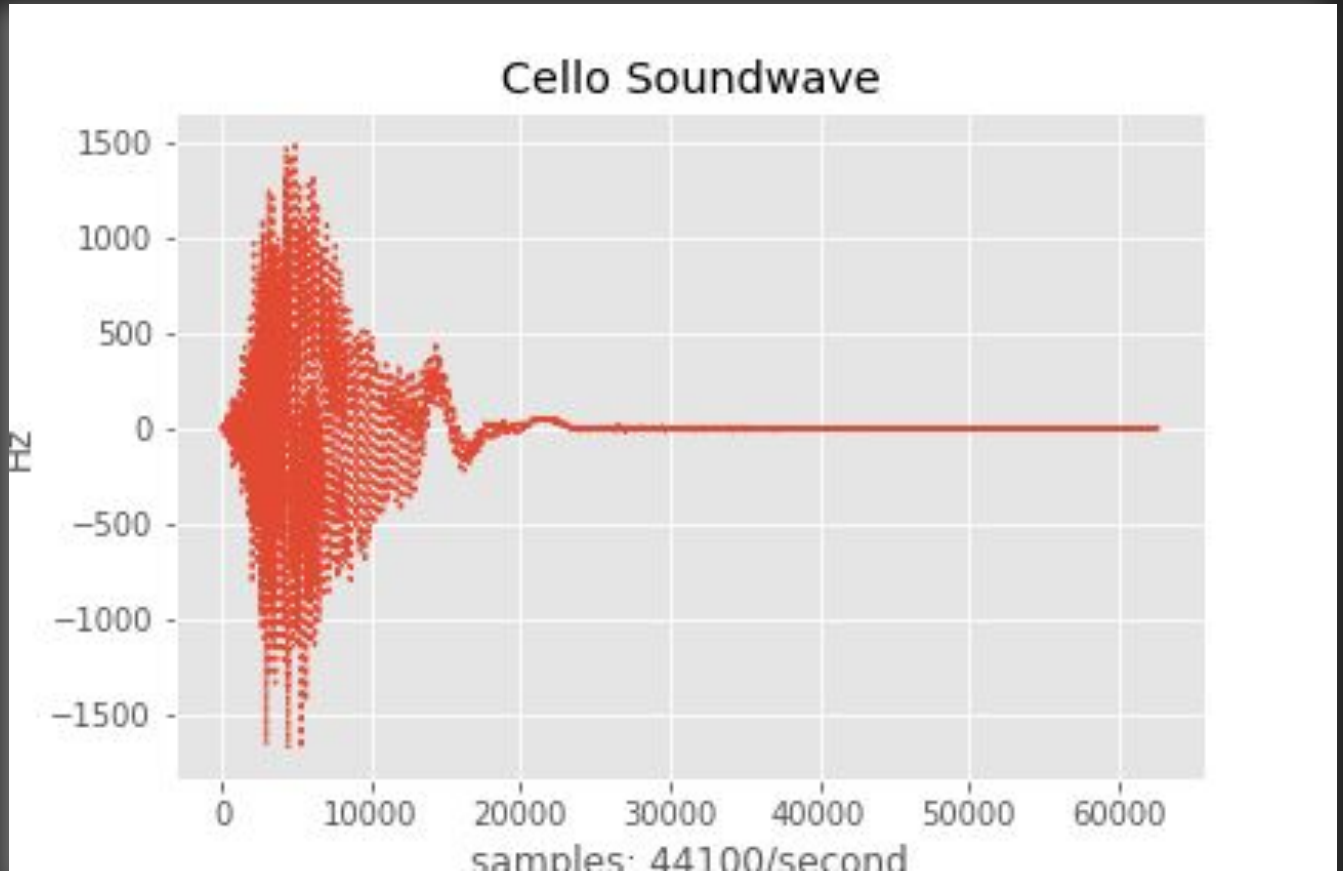
Plug-in to Signal Processing

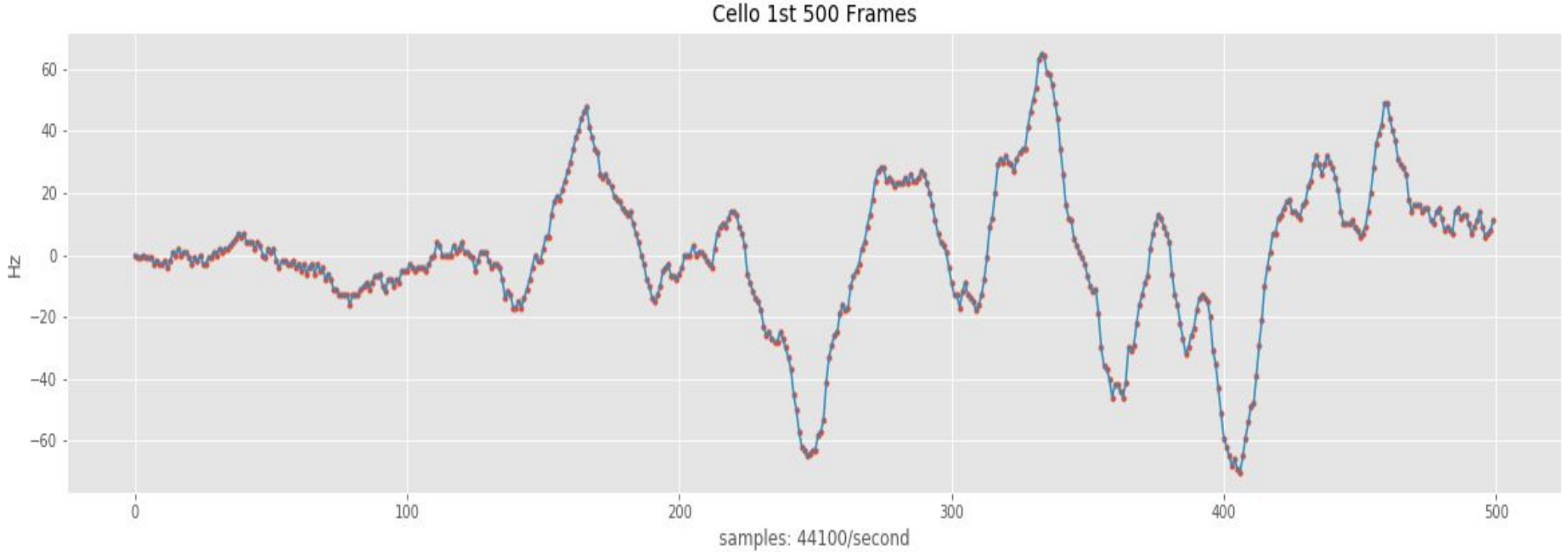
Making sense of a audio world

Sound is Vibration

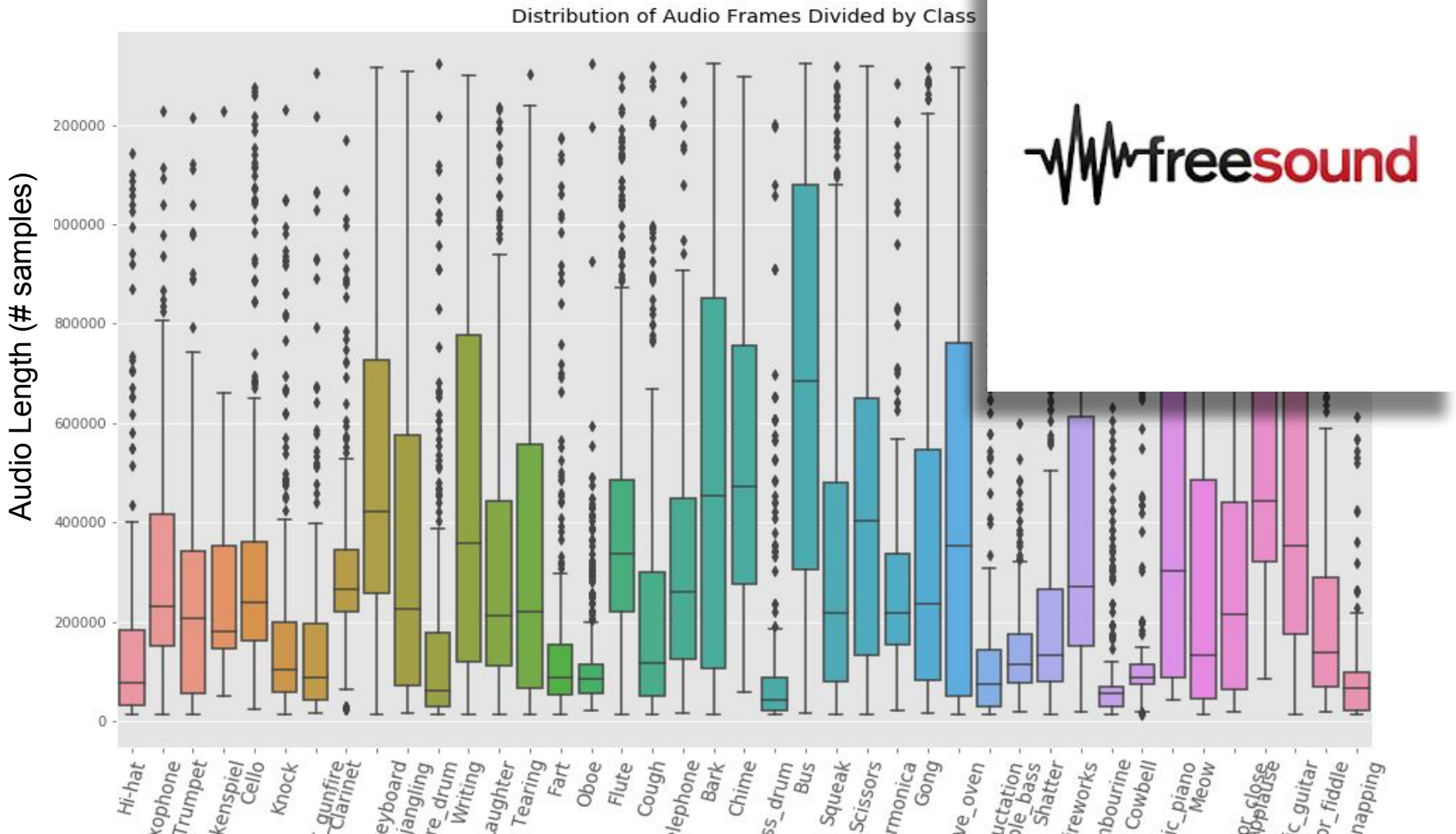


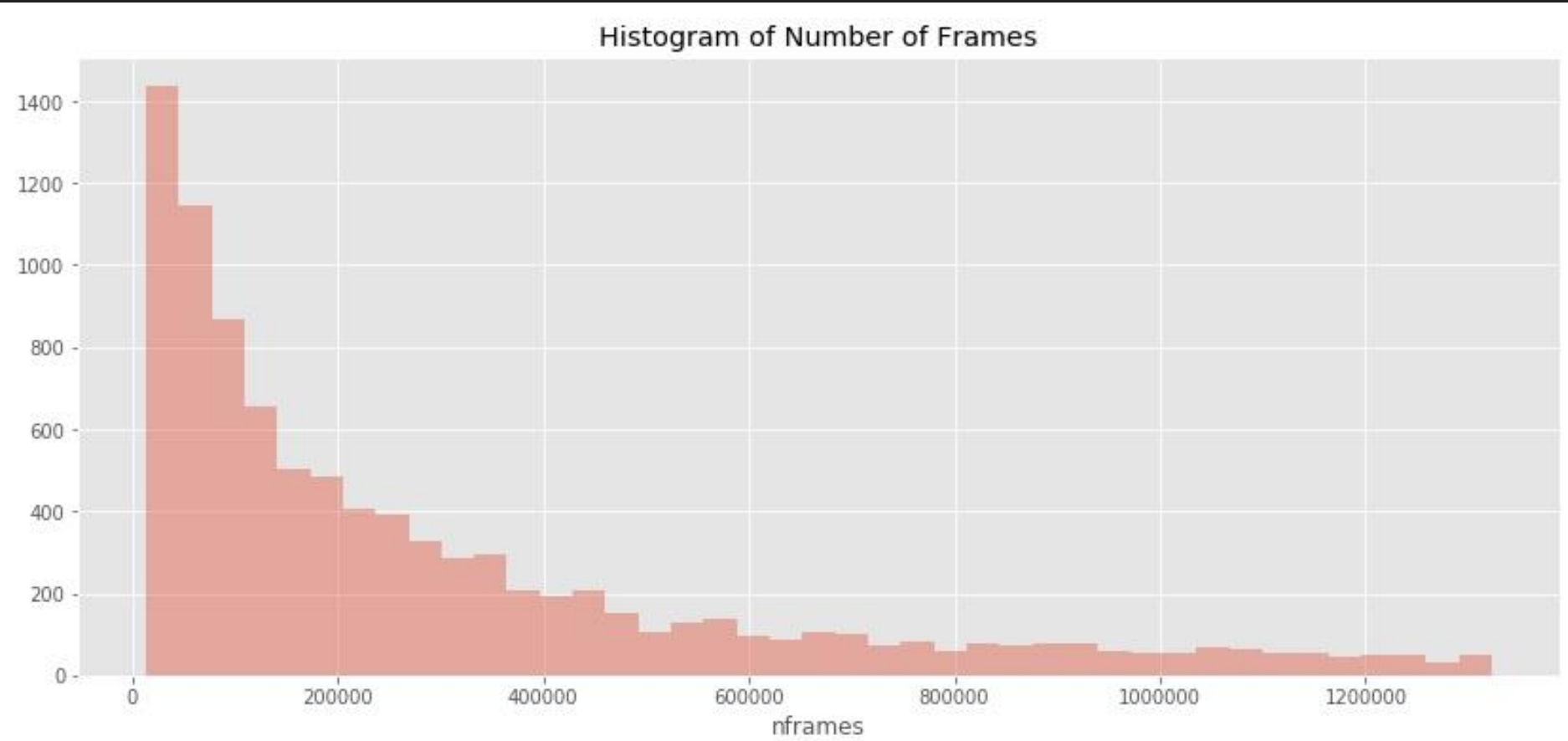
Sound is Vibration




Cello 1st 500 Frames
Hz
60
40
20
0
−20
−40
−60
0
100
200
300
400
500
samples: 44100/second

CNN or RNN

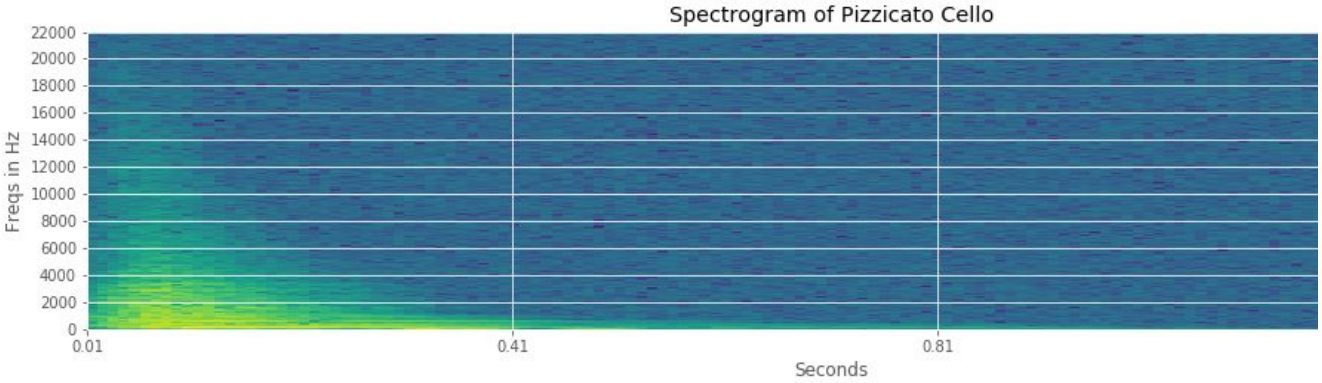

Distribution of Audio Frames Divided by Class
Audio Length (# samples)
200000
000000
800000
600000
400000
200000
0
Hi-hat
Trumpet
Cello
Knock
Clarinet
Writing
Tearing
Fart
Oboe
Flute
Cough
Bark
Chime
Bus
Squeak
Scissors
Gong
Cowbell
Meow
Shatter
Tearing
Fart
Oboe
Flute
Cough
Bark
Chime
Bus
Squeak
Scissors
Gong
Cowbell
Meow
Shatter
freesound

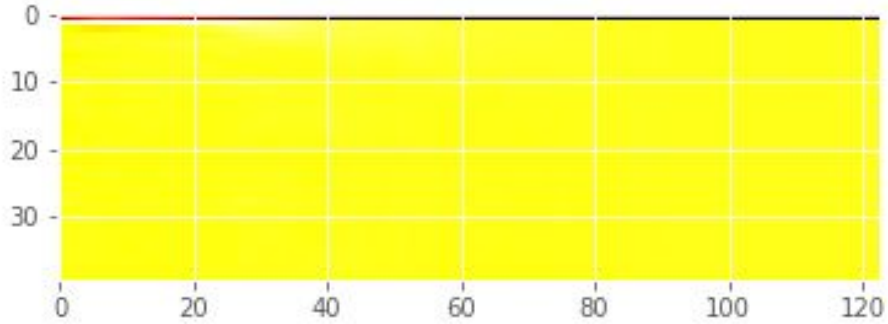

Histogram of Number of Frames
1400
1200
1000
800
600
400
200
0
0
200000
400000
600000
800000
1000000
1200000
nframes

Composing

1. Join audio files on labels
2. Preform data augmentation

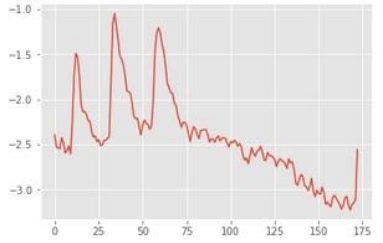
Data Augmentation





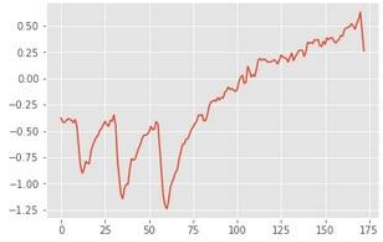

```
[69]: plt.plot(MFCC_X_train[0][2])
```

```
[69]: [<matplotlib.lines.Line2D at 0x26b42d56208>]
```



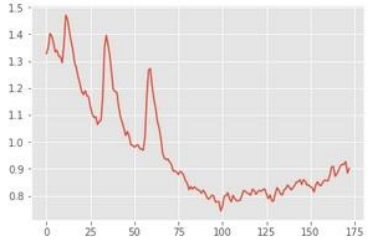

```
[70]: plt.plot(MFCC_X_train[0][1])
```

```
[70]: [<matplotlib.lines.Line2D at 0x26b42dae668>]
```




```
[71]: plt.plot(MFCC_X_train[0][0])
```

```
[71]: [<matplotlib.lines.Line2D at 0x26b42e14940>]
```




```
[73]: MFCC_X_train[0].shape
```

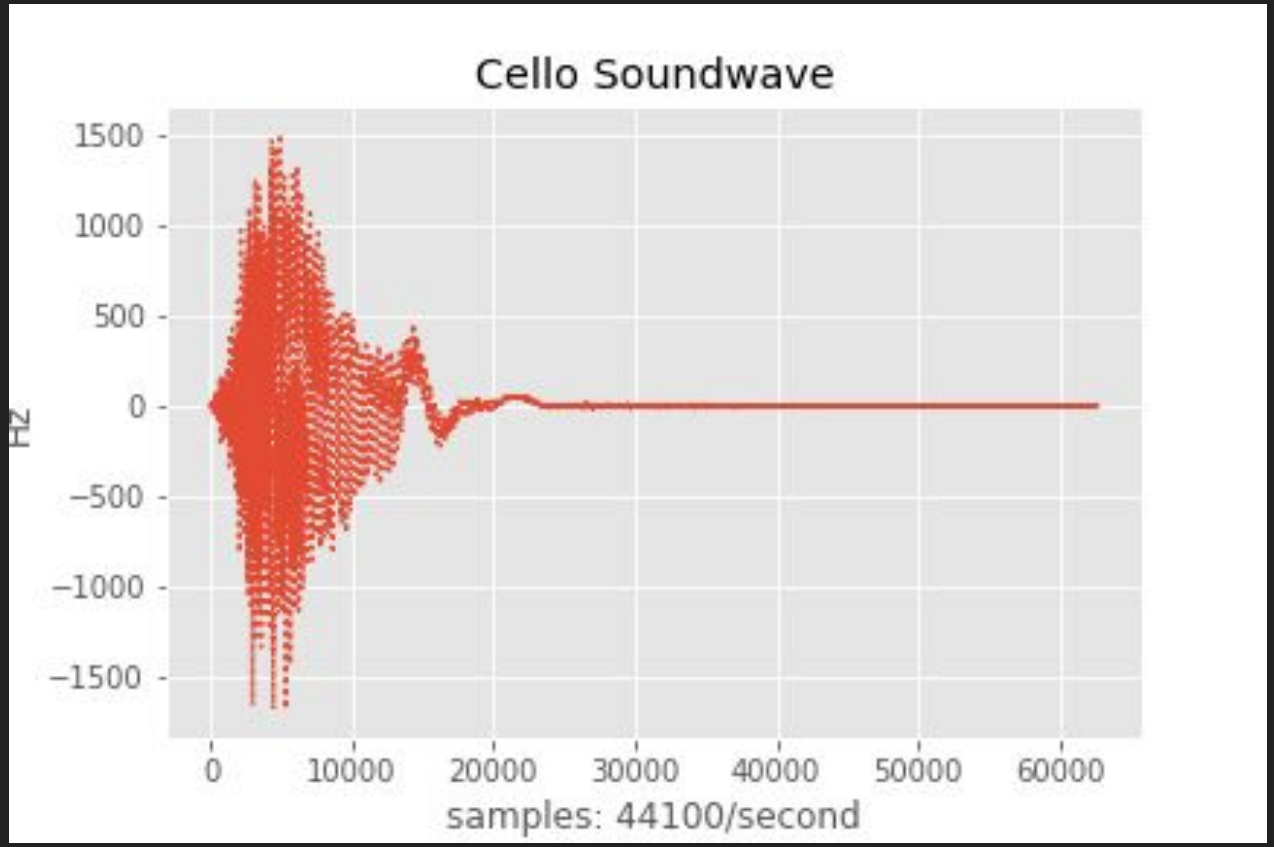
```
[73]: (40, 173, 1)
```

Composing

3. Clean data (windowing & signal from noise)

4. Train model

Cropping Audio



1D CNN

```
Fold:  0
#####
Epoch 1/1
14/14 [==============================] - 133s 9s/step - loss: 3.7228 - acc: 0.0078 - val_loss: 3.7255 - val_acc: 0.0132

Epoch 00001: val_loss improved from inf to 3.72551, saving model to best_0.h5
Fold:  1
#####
Epoch 1/1
15/15 [==============================] - 145s 10s/step - loss: 3.7226 - acc: 0.0219 - val_loss: 3.7186 - val_acc: 0.0270

Epoch 00001: val_loss improved from inf to 3.71864, saving model to best_1.h5
```

2D CNN

```
#####
Fold:  0
Train on 1781 samples, validate on 219 samples
Epoch 1/2
1781/1781 [==============================] - 65s 36ms/step - loss: 3.5222 - acc: 0.1078 - val_loss: 3.2974 - val_acc: 0.1644

Epoch 00001: val_loss improved from inf to 3.29743, saving model to best_0.h5
Epoch 2/2
1781/1781 [==============================] - 65s 36ms/step - loss: 2.9545 - acc: 0.2212 - val_loss: 3.0789 - val_acc: 0.1918

Epoch 00002: val_loss improved from 3.29743 to 3.07893, saving model to best_0.h5
2000/2000 [==============================] - 17s 9ms/step
```

Further Reading...

https://www.kaggle.com/fizzbuzz/beginner-s-guide-to-audio-data : audio classification

https://towardsdatascience.com/recognizing-speech-commands-using-recurrent-neural-networks-with-attention-c2b2ba17c837 : RNN for speech recognition

https://ai.google/research/teams/brain/magenta : generative audio model

https://www.google.com/doodles/celebrating-johann-sebastian-bach : Google doodle using magenta

http://coding-geek.com/how-shazam-works/ : how Shazam works!!

https://www.linkedin.com/in/mark-ehler-85052548/ : me