SPECTROGRAM ANALYSIS
FOR INSTRUMENT PITCH COMPARISON

SERGIO LATROFA - 640584

ANALYSED **INSTRUMENTS**

TWO INSTRUMENTS FOR TWO "FAMILIES"

a **Violin family** (strings)
Sound waves are produced by a **bow** rubbing strings.

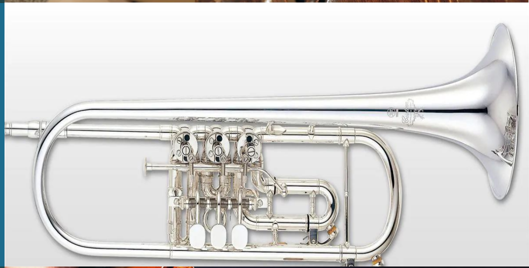
b **Brass family** (wind)
Different pitches are produced using **slides**, **keys** and **crooks**.

Trumpet and **violin** are expected to produce *higher* sound with respect to **cello** and **trombone**, whose pitch is considerable *lower*.
So similarity may be detected both between instruments belonging the same *family* rather than ones closer in terms of sound *height*.

When plotting spectrograms, each column will correspond to an instrument: (1) Cello, (2) Violin, (3) Trombone, (4) Trumpet.

Cello
Largest violin after contrabassoon. Played on the ground. Low sound.

Trumpet
Small brass instrument equipped with keys. Sound range is quite high..



Violin
The smallest of its family. It's played on the shoulder. Sound is high..

Trombone
Equipped with a large sliding *coulisse*.
Several sizes exist, with generally middle-low sound.

CHOSEN **NOTE SETS**

ISOLATE VARIATING FACTORS (DYNAMIC & HEIGHT)

A huge constraint was the non strict regularity in the dataset: basing on note ids, not every dynamic was present for each note.

For arc instruments, only "*arco normal*" played notes were picked to avoid introducing another source of complexity.

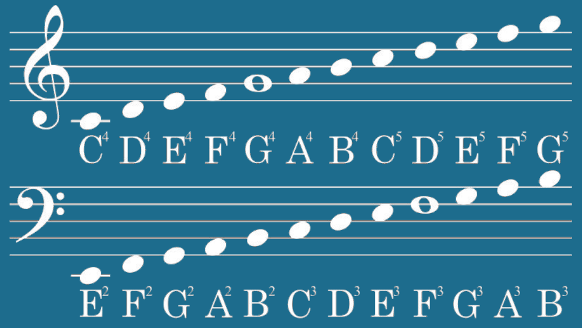
Each set aims to expose how different aspects impacts on waves and spectrum .

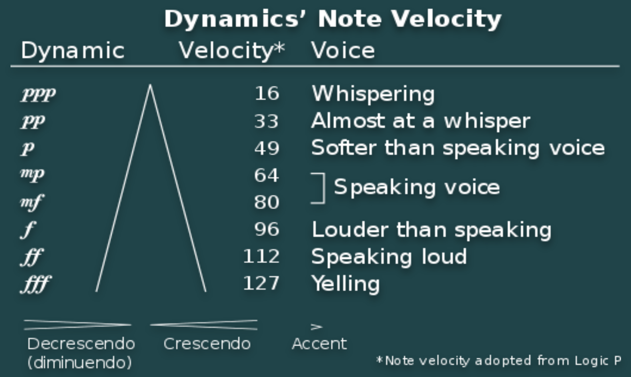
`{A3_1, B4_025,G5_15}` `{C4_1, C5_1, C6_1}` `{B4_05} X {pp,f,ff}`

Random notes
Same dynamic

Different octaves
Same dynamic and note

Different dynamics
Same note





Dynamics' Note Velocity

Dynamic	Velocity*	Voice
ppp	16	Whispering
pp	33	Almost at a whisper
p	49	Softer than speaking voice
mp	64	Speaking voice
mf	80	Speaking voice
f	96	Louder than speaking
ff	112	Speaking loud
fff	127	Yelling

Decrescendo (diminuendo) Crescendo Accent

*Note velocity adopted from Logic P

ABOUT **METHOD** AND **CODE**

Librosa and *matplotlib* are the hinges of this experimentation.

Documentation examples were followed to implement a the various plotting functions considerable useful for the analysis.

A global high order function was implemented to display subplots in a grid and make observation easier.

Harmonics isolation was also implemented ,as an option, within this function, allowing visualize difference with respect to the full spectrum.

The full notebook can be viewed just by clicking on the Colab Icon in this slide.

```
#Compute STFT to visualize spectrogram
def full_spectrogram(y, sr, title="", plot=plt.subplots()):
    D= librosa.stft(y)
    rp = np.max(np.abs(D))
    librosa.amplitude_to_db(np.abs(D), ref=rp)
    fig, ax = plot
    img = librosa.display.specshow(D, ax=ax,
                                   y_axis='log', x_axis='time')
    ax.set(title=title + ': Power Spectrogram')
    ax.label_outer()
    fig.colorbar(img, ax=ax)
    return D, 'stft'
```

```
# Wrapping plotting function to display all images closer
nnotes = 3
ninstruments = 4

def grid_plot(plotting_func, h_only=False, save_feature=True):
  fig, axarr  = plt.subplots(nrows=nnotes, ncols=ninstruments)
  fig.set_size_inches(10*ninstruments, 6*nnotes)

  # Dislpay plots having notes on rows and instruments on columns
  for i, sound in enumerate(notes):
    row = i % nnotes
    col = i // nnotes
    name, note = sound

    # Analyzing only harmonics can be useful
    if(h_only):
      note = (librosa.effects.hpss(note[0])[0], note[1])

    #Execute transformation, features are saved for further usage.
    feature, feature_name = plotting_func(note[0], note[1], name, (fig, axarr[row, col] ))

    #Harmonics != Full
    if(h_only):
      feature_name += '_harmonics'

    # Saving computed stuff for further usage
    if(save_feature):
      features[name][feature_name] = feature

  plt.tight_layout()
```

Preprocessing

All considered notes were uploaded on a new Google Drive folder and loaded into the Colab Notebook using librosa.load function.

Each extracted time-series was then trimmed, applying again a librosa function to delete “silence” moment (according to a predetermined threshold).

COMPUTATION: **STFT**

librosa compute spectrograms using the **Short Time (Discrete) Fourier Transform**.

The **STFT** represents a signal in the time-frequency domain by computing discrete Fourier transforms (**DFT**) over short overlapping windows.

The number of rows in the STFT matrix **D** is (*1 + n_fft/2*). The default value, *n_fft=2048* samples, corresponds to a physical duration of *93 milliseconds* at a sample rate of *22050 Hz,* i.e. the default sample rate in *librosa*.

Theoretical foundation

There is no significant information lost if we replace the short term spectrum ***X(f, t)*** with its sampled version ***Xnm*** by sampling it at the two *Nyquist* periods **1/F** and **1/T**.

$$X(f,t) = \int_{-\infty}^{\infty} w(t-\tau)x(\tau)e^{j2\pi f\tau}d\tau$$

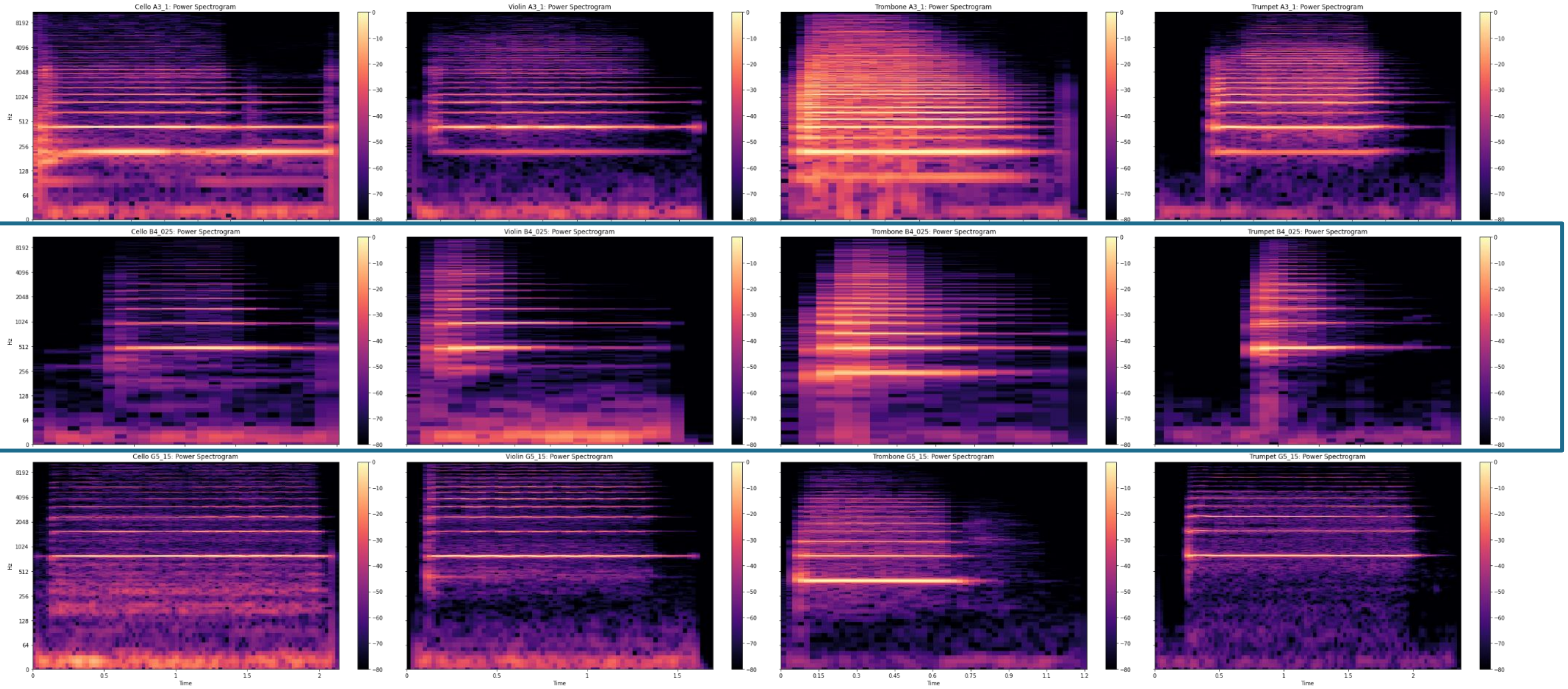
$$W(f) = \int_{-\infty}^{\infty} w(\tau)e^{j2\pi f\tau}d\tau$$

$$X_{nm} = \sum_{k=0}^{T-1} w(nD-k)x(k)e^{j2\pi km/T}$$

$$D = 1/F$$

1 Jont B. Allen (June 1977). **"Short Time Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform".** *IEEE Transactions on Acoustics, Speech, and Signal Processing*. ASSP-25 (3): 235–238.

FULL POWER SPECTROGRAM



RANDOM NOTES - Spectre intensity is similar within rows, repetition seems to occur in among the horizontal bands (harmonics).

COMPUTATION: **HARMONICS EXTRACTION**

Minimization the L2 norm of the power spectrogram gradients allows, formally, to achieve separation.

Median filters operates replacing a given sample in a signal by the median of the signal values in a window around the sample, smoothing both the percussive and harmonics components. .

$$J(\mathbf{H},\mathbf{P}) = \frac{1}{2\sigma_H^2}\sum_{h,i}(H_{h,i-1} - H_{h,i})^2 + \frac{1}{2\sigma_P^2}\sum_{h,i}(P_{h-1,i} - P_{h,i})^2$$

$$y(n) = \mathrm{median}\{x(n-k : n+k), k = (l-1)/2\}$$

$$P_i = \mathcal{M}\{S_i, l_{\mathbf{perc}}\}$$

$$H_i = \mathcal{M}\{S_h, l_{harm}\}$$

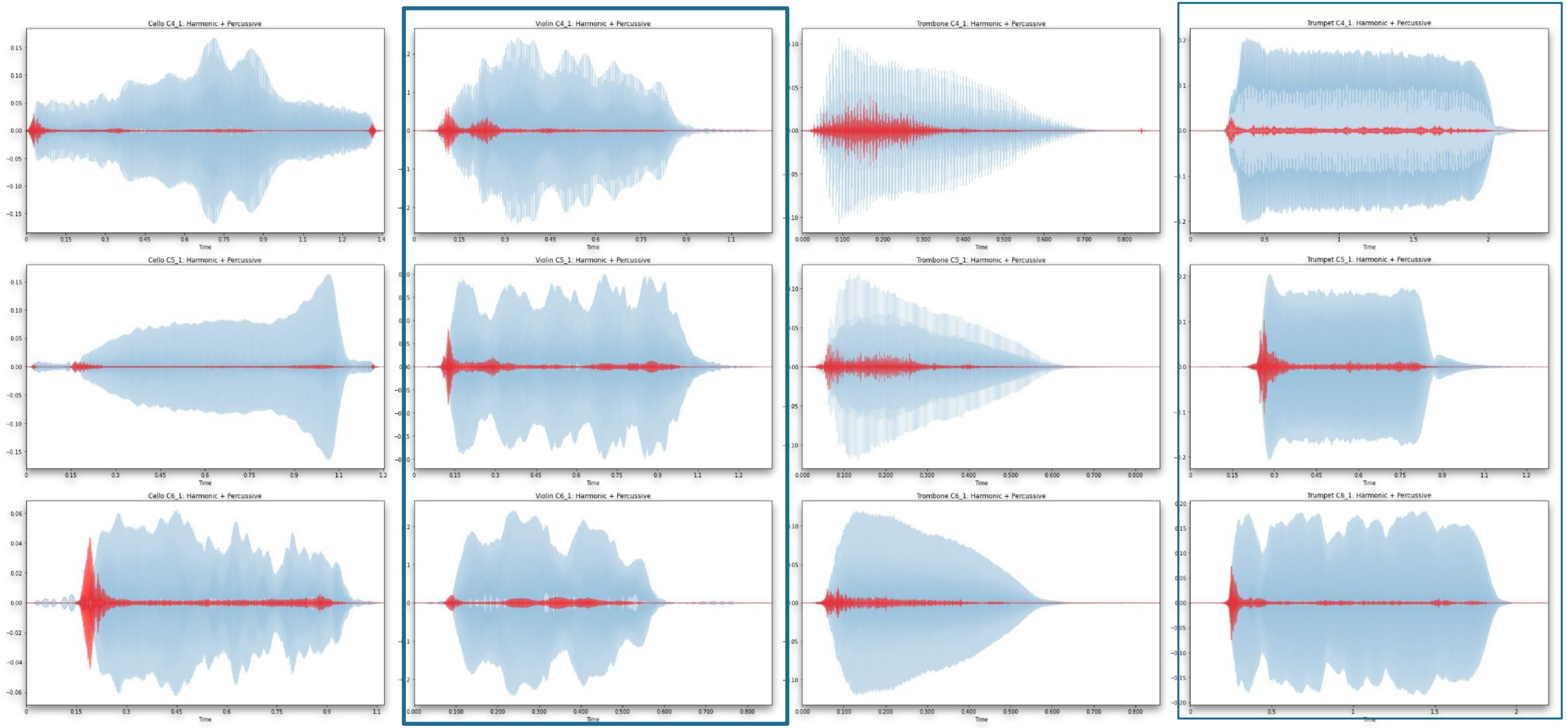
Intuition behind

This technique was based on the idea that stable harmonic or stationary components form horizontal ridges on the spectrogram, while percussive components form vertical ridges with a broadband frequency response.

1 Fitzgerald, Derry. **“Harmonic/percussive separation using median filtering.”** 13th International Conference on Digital Audio Effects (DAFX10), Graz, Austria, 2010.

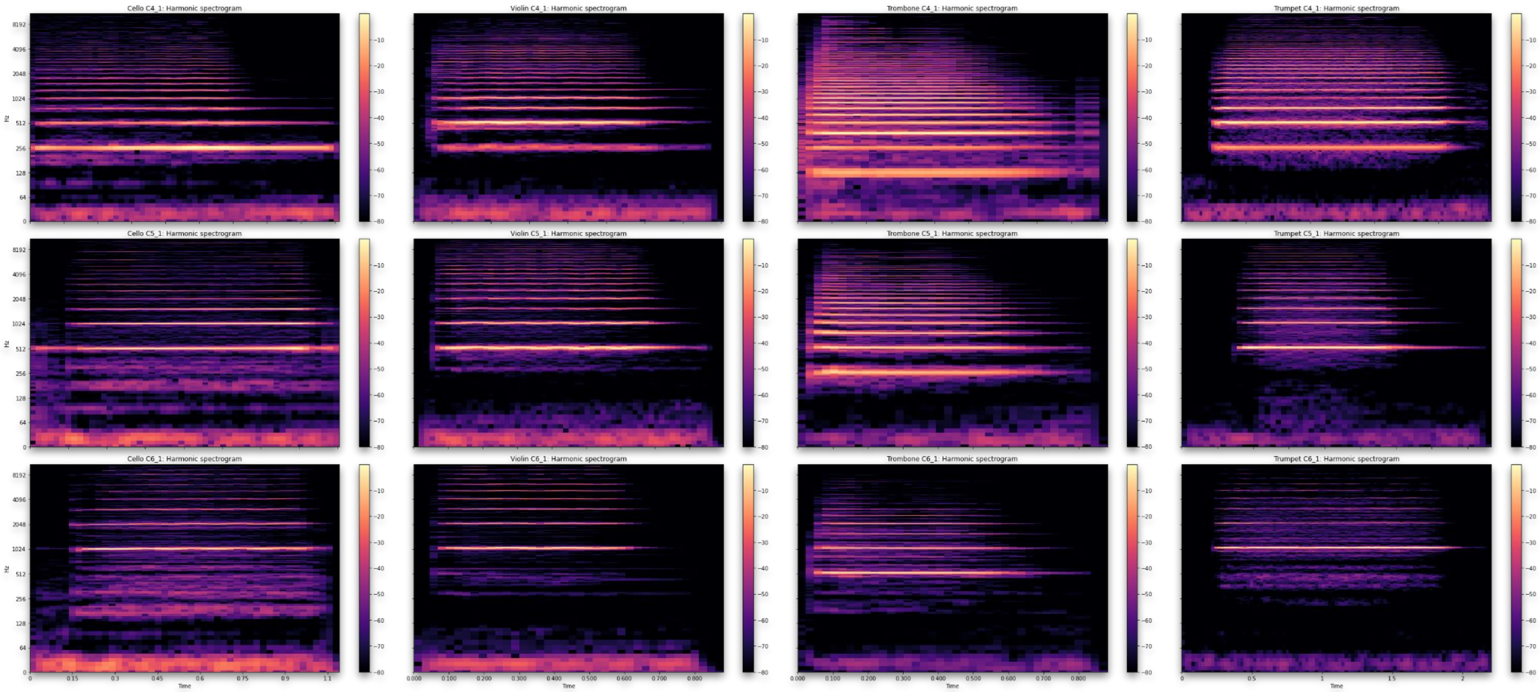
2 Driedger, Müller, Disch. **“Extending harmonic-percussive separation of audio.”** 15th International Society for Music Information Retrieval Conference (ISMIR 2014), Taipei, Taiwan, 2014.

HARMONICS PERCUSSIVE **DECOMPOSITION**



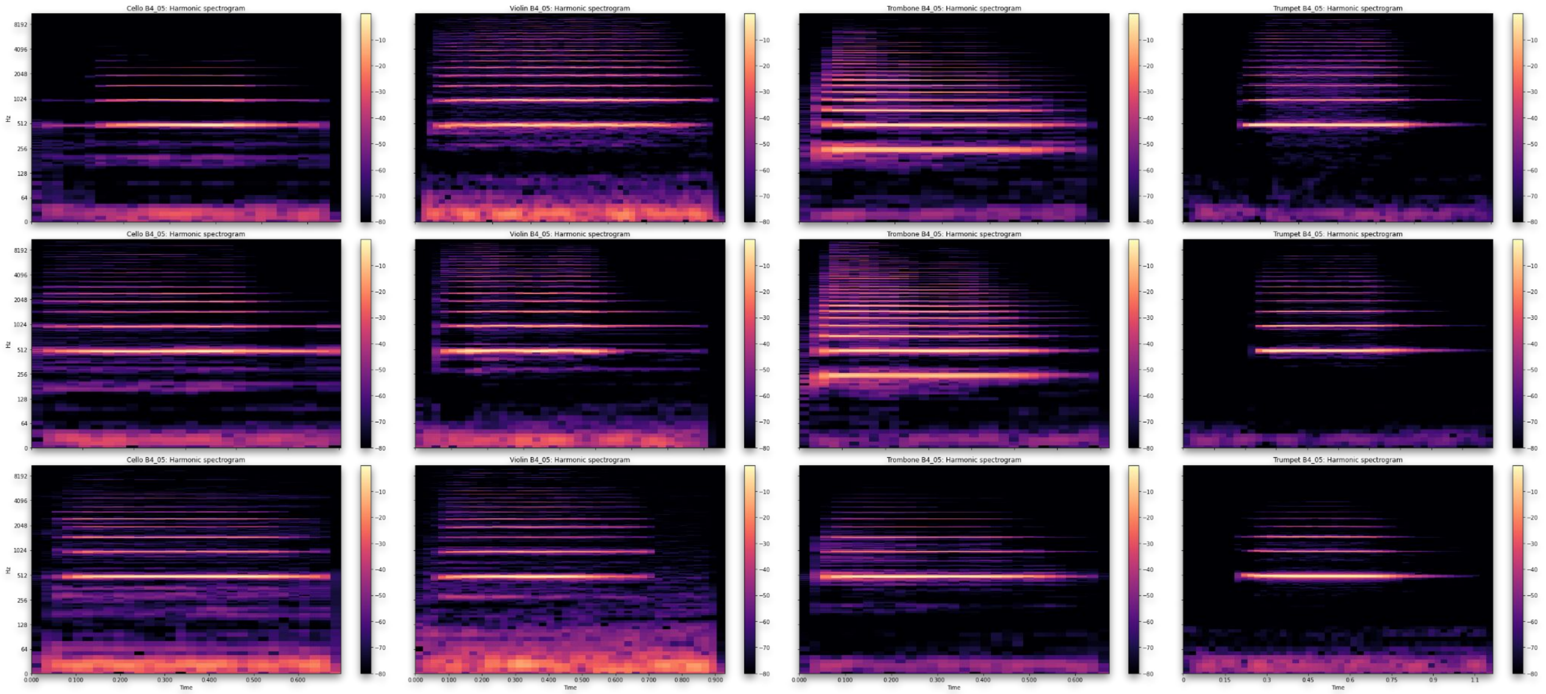
SAME NOTE AT DIFFERENT OCTAVES - Shapes of harmonics(blue) is similar among instrument,especially for violin and trumpet.

HARMONICS: **SIMILAR SPECTROGRAMS**



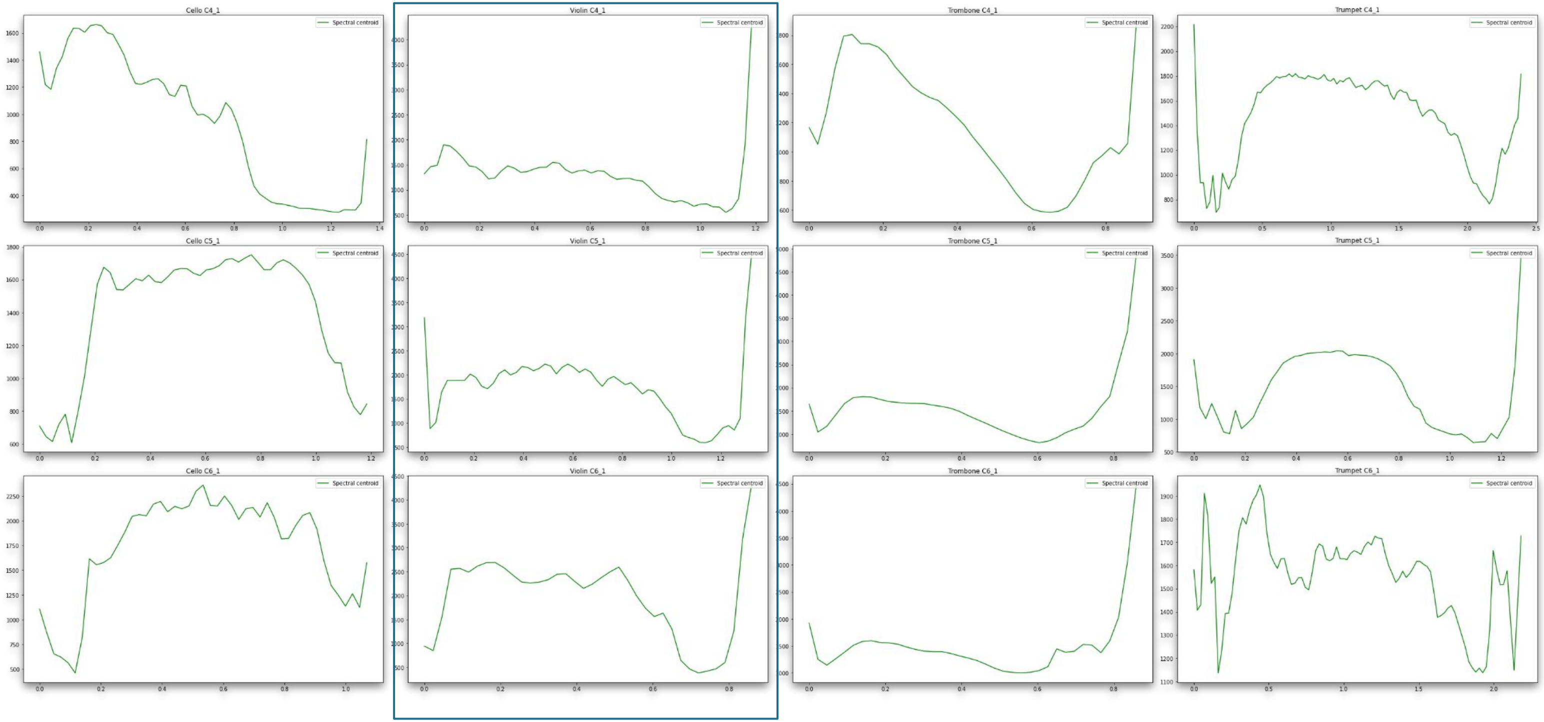
SAME NOTE AT DIFFERENT OCTAVES - Spectre intensity change within column, but intensity levels seems to be repeating.

HARMONICS: **SIMILAR SPECTROGRAMS**



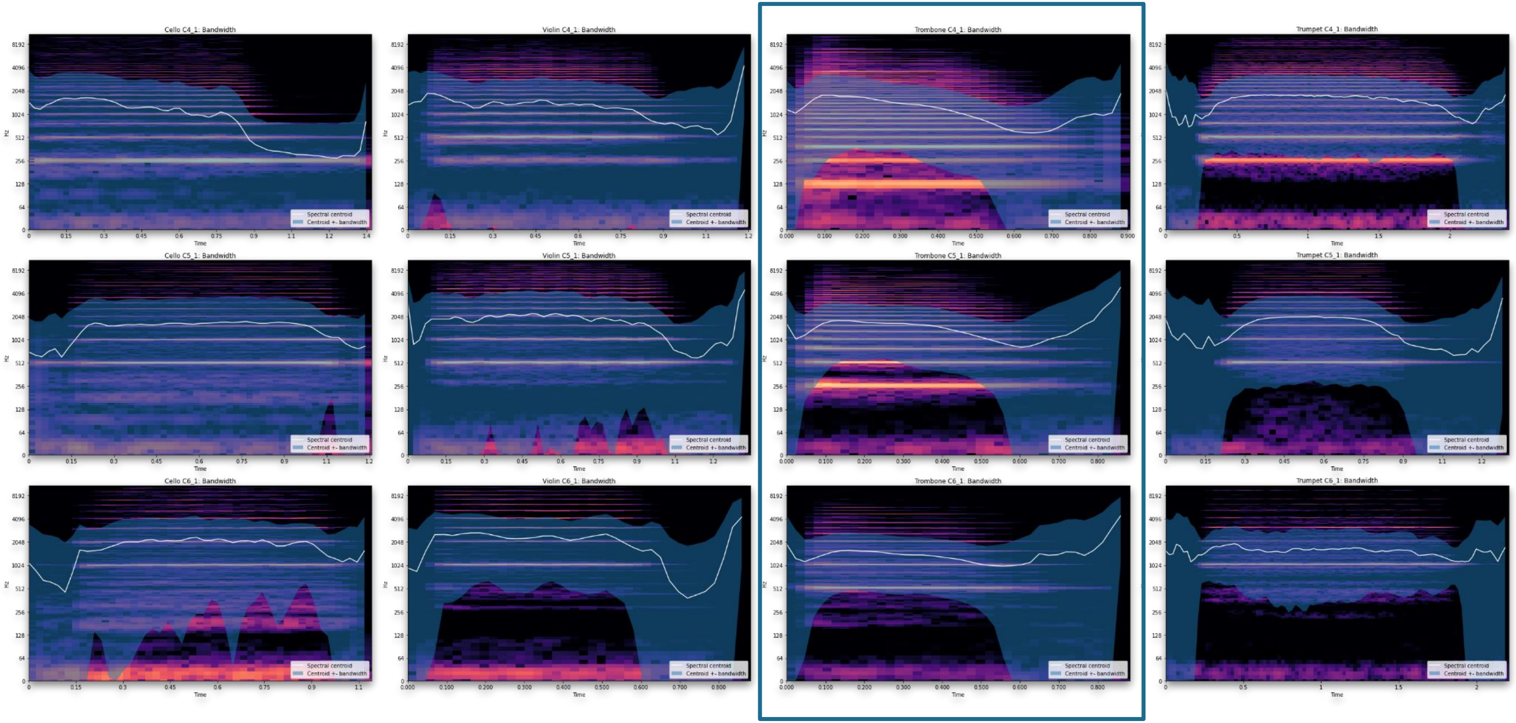
SAME NOTE WITH DIFFERENT DYNAMICS - Similarities among frequency bands is evident, intensity differs much.

HARMONICS: **SPECTRAL CENTROIDS**



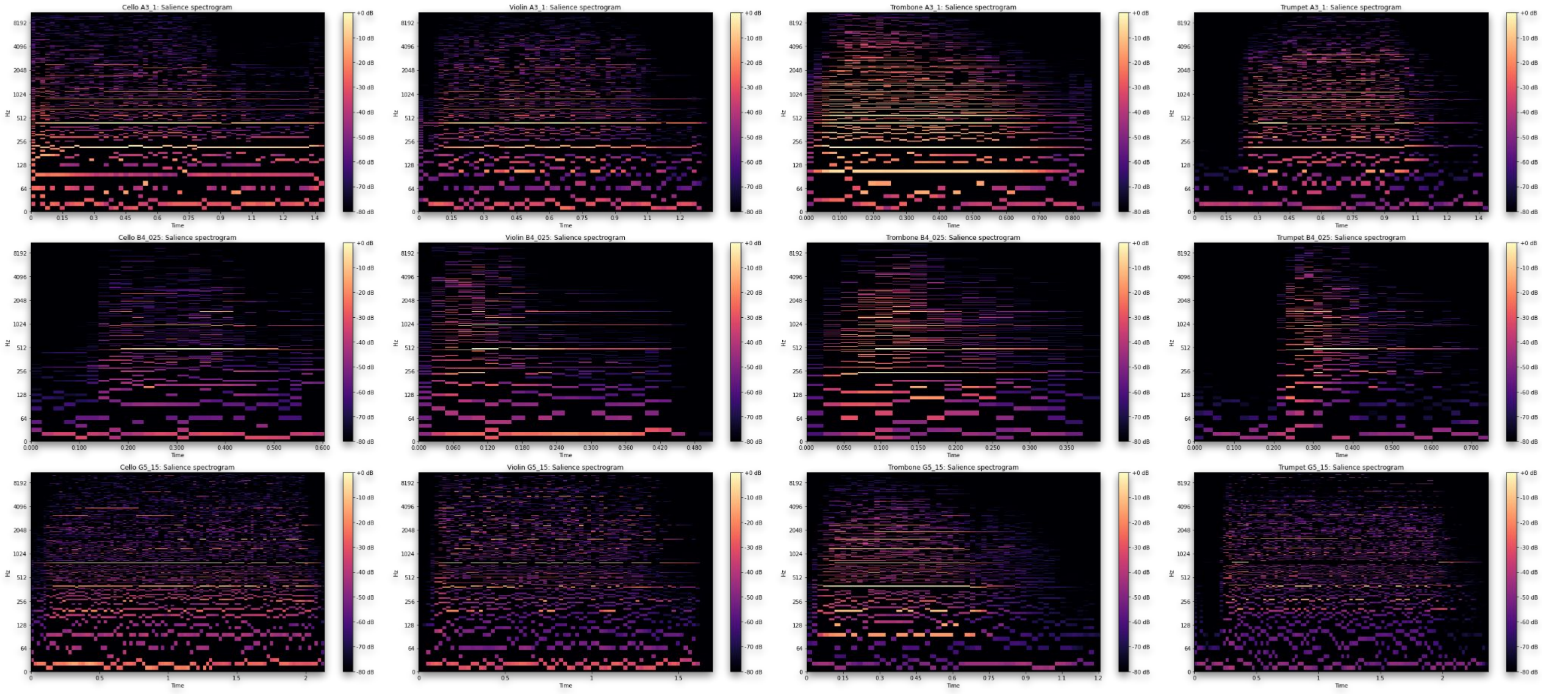
SAME NOTE AT DIFFERENT OCTAVES - Behaviour of centroids is similar within columns, average frequence reflect note “height”.

BANDWIDTH AND CENTROIDS



SAME NOTE AT DIFFERENT OCTAVES - Clear similarity within same columns (same instrument), especially for Trombone (col 3.)

VISUALIZATION: **SALIENCE SPECTROGRAM**



RANDOM NOTES - Salience plots highlight more similarity among notes than among instruments.

FINAL CONSIDERATION

ABOUT INSTRUMENT PITCH RECOGNITION

Each instrument shown a different **harmonic profile**. Instrument within the **same family** have **closer profiles**.

Fixing a note helped a lot reducing complexing, highlighting key similarities and differences.
Octave change produced a **vertical translation** of the harmonic profile, following the logarithmic scale.

Dynamic variation was translated into a (proportional) **decreasement of energy level** for the various harmonics, making spectre of various note similar even if played by different instruments.

FURTHER "WEIRD" IMPROVEMENTS

Several ideas occurred while performing the assignment, like the creation of simplified feature (mean, std, min, max etc..) from computed spectrograms and representation, allowing to computes simple DM algorithms like K-means or even PCA on note-vectors.

Papers on MIR using CNN already exist, maybe they would be useful for the Deep Learning assignment.

ABOUT THE ASSIGNMENT

Theoretical deepening

Concept clarified reading papers and experimenting their direct application.

New skills as a developer

Experience with librosa and matplotlib for time-series and spectrograms.

Advanced tools explored

STFT, HPPS decomposition, spectral centroids and bandwidth, salience and more...