MUSIC INFORMATION CLASSIFICATION

PROGRESS PRESENTATION BY

M.BERKAN ERCAN

A SHORT REMINDER

Music Information Retrieval (MIR), is related with extracting information from music data.

The data can be;

- Genre
- Timbre
- Lyrics
- Mood
- Pitch Sounds
- Tempo…

THE PROJECT

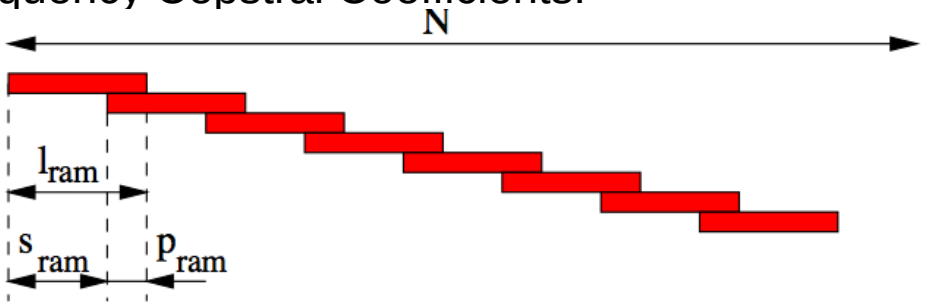
- **The aim was to create a map of meodies around Turkey.**
- **To match the given composer and region information to the songs in the training dataset. (1300 Turkish folk songs)**
  - Find unique representation of the melodies
  - So the proposed system will be able to identify a novel melody’s origin.

THE FIRST CHALLENGE

Feature Extraction is the first step in an MIR system.

One of the possible methods;

- MFCC
  - Defining frames on the music data and extracting Mel-Frequency Cepstral Coefficients.



$$N_{ram} = 1 + \left\lfloor \frac{N - l_{ram}}{s_{ram}} \right\rfloor$$

  - In order to perform MFCC, first we have use FFT.
  - Next, the results are compared with the Mel-scale.

THE FIRST CHALLENGE (CONT.)

The other possible way is to divide the melody into chunks and appy FTT directly.

- **Commercial products working this way(Shazam)**

THE PROGRESS

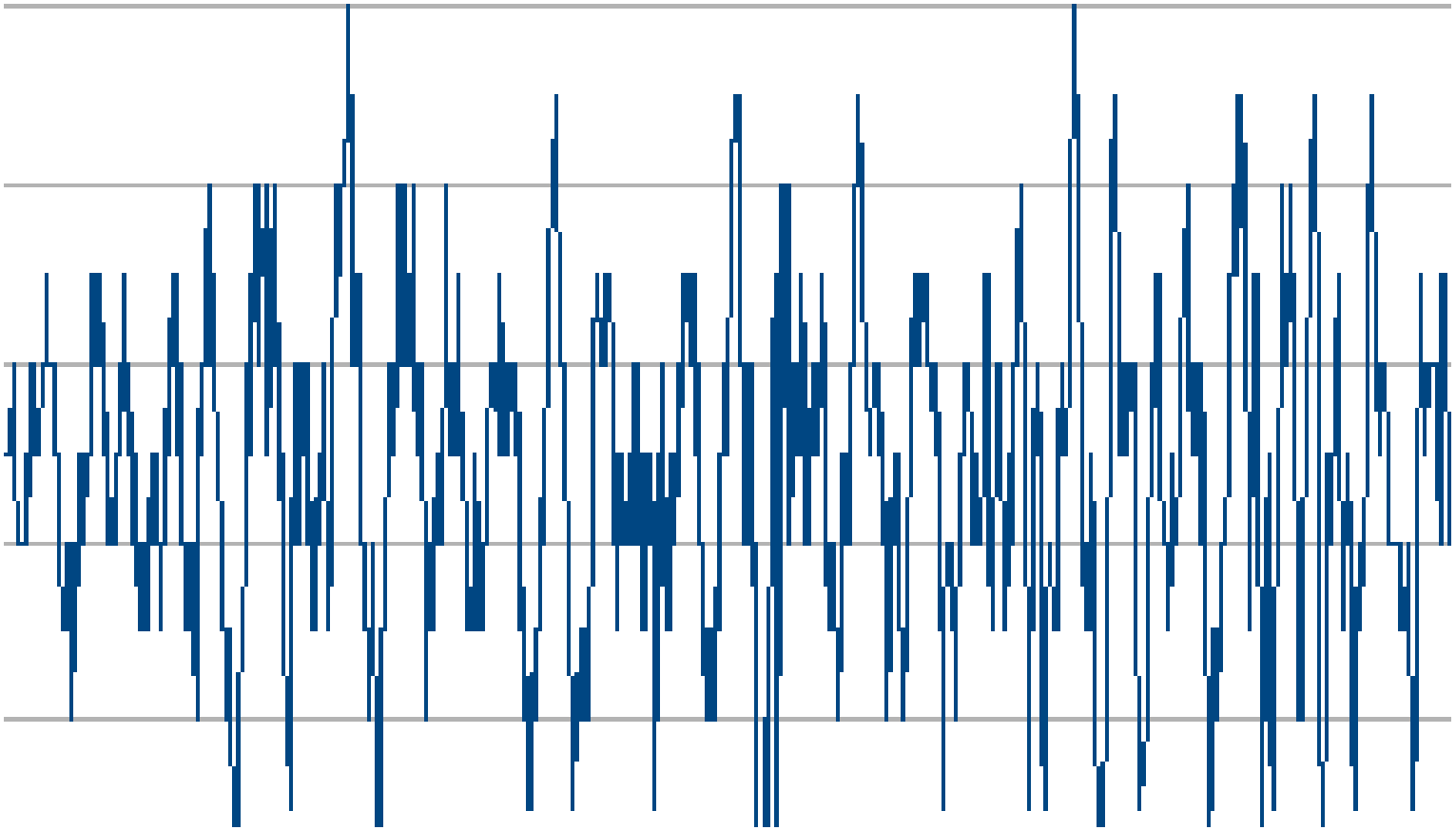
The progress can be categorized as follows;

- Get the audio file from microphone (10 sec)
- Divide it into chunks of 4096 bytes
- Apply FFT to each chunk in order to switch from time-domain to frequency domain.
- Create a unique representation from the frequency distribution of the melody.

AUDIO DATA IN TIME DOMAIN

```
-1 -1 -1 -1 -1 -1 0 0 0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 4 5 5 5 5 5 5 5 6 6 6 6 6 6
5 5 5 5 5 5 5 4 4 4 4 3 3 3 3 2 2 2 1 1 0 0 -1 -1 -2 -2 -3 -4 -4 -5 -6 -6 -7 -7 -8 -9 -9
-10 -10 -10 -11 -11 -12 -12 -12 -12 -13 -13 -13 -13 -13 -14 -14 -14 -14 -14 -14 -14 -14
-14 -15 -15 -15 -15 -15 -15 -14 -14 -14 -14 -14 -14 -13 -13 -13 -12 -12 -12 -11 -11 -10
-10 -10 -9 -9 -8 -8 -8 -8 -7 -7 -7 -7 -6 -6 -6 -6 -5 -5 -5 -5 -4 -4 -4 -3 -3 -3 -2 -2 -1
-1 -1 0 0 1 1 1 2 2 3 3 3 3 4 4 4 5 5 5 5 6 6 6 6 7 7 7 8 8 8 8 9 9 9 10 10 10 10 11 11 11
11 11 11 11 12 12 12 12 12 12 12 12 12 12 12 12 11 11 11 11 11 11 11 11 11 11 10 10 10 10
10 10 9 9 9 9 8 8 8 8 7 7 7 6 6 6 5 5 5 5 4 4 4 3 3 3 2 2 2 1 1 1 1 0 0 0 -1 -1 -1 -1 -2
-2 -2 -2 -3 -3 -3 -3 -4 -4 -4 -5 -5 -5 -5 -6 -6 -6 -6 -7 -7 -7 -7 -7 -7 -8 -8 -8 -8 -8 -8
-8 -9 -9 -9 -9 -9 -9 -9 -10 -10 -10 -10 -10 -10 -10 -10 -10 -10 -11 -11 -11 -10 -10 -10
-10 -10 -10 -10 -10 -10 -9 -9 -9 -9 -9 -8 -8 -8 -8 -7 -7 -7 -6 -6 -6 -5 -5 -5 -4 -4 -3 -3
-2 -2 -1 -1 -1 0 0 1 1 2 2 3 3 4 4 5 5 6 6 6 7 7 8 8 8 9 9 9 10 10 10 10 11 11 11 11 11 12
12 12 12 12 12 12 12 12 12 12 11 11 11 11 10 10 10 9 9 8 8 7 7 6 6 5 5 4 3 3 2 1 1 0 -1 -1
-2 -3 -4 -4 -5 -6 -7 -8 -9 -10 -10 -11 -12 -13 -13 -14 -15 -15 -16 -16 -17 -17 -17 -18 -18
-18 -18 -18 -19 -19 -19 -19 -19 -19 -19 -19 -19 -20 -20 -20 -20 -20 -20 -20 -19 -19 -19
-19 -18 -18 -18 -17 -17 -16 -16 -15 -15 -14 -14 -13 -13 -12 -12 -12 -11 -11 -10 -10 -10 -9
-9 -9 -8 -8 -8 -7 -7 -7 -6 -6 -5 -5 -4 -4 -3 -3 -2 -1 -1 0 0 1 2 2 3 3 4 4 4 5 5 6 6 6 7 7
7 8 8 8 9 9 9 10 10 11 11 11 12 12 12 12 13 13 13 13 14 14 14 14 14 14 14 14 14 14 14 15
14 14 14 14 14 14 14 14 14 14 14 14 14 14 13 13 13 13 13 13 12 12 12 12 11 11 11 11 10 10
10 9 9 9 8 8 8 7 7 7 6 6 5 5 5 4 4 3 3 3 2 2 1 1 1 0 0 -1 -1 -1 -2 -2 -2 -3 -3 -3 -4 -4 -5
-5 -5 -6 -6 -7 -7 -7 -8 -8 -9 -9 -9 -10 -10 -10 -11 -11 -11 -11 -11 -12 -12 -12 -12 -12
-12 -12 -12 -12 -12 -12 -13 -13 -13 -12 -12 -12 -12 -12 -12 -12 -12 -11 -11 -11 -11 -10
```

AUDIO DATA IN TIME DOMAIN



AUDIO DATA IN THE FREQUENCY DOMAIN

948+-2.6538190535052131i5.6124675789582525+-3.44225191499112351i-15.962387688551187+-16.6520
45145955984i-13.6442644783633375+8.1145136850861718i8.022439351009031+15.923380059188081i1.5591
326136245804+22.6853987030398513.997279228103669+23.597861194463351-0.601855269247594+-3.6
029350910993471+4.28603732971092+4.847583938071533513.45190904073016318.689348114980431-8.9
4970421946601+3.3440580221847496i-10.8034917075933335.3148663990346241+2.08845928638494451+
-12.78837024702658i-4.278293073906549+-10.9101311899795311.872605354119342+5.0829851269153
821-3.0018294335187403+-3.060584574324885i-7.0132671784410711+-8.4486953165072731-10.122371
174000055+-11.405581447931004i-13.244519408330543+2.3469117824634651-3.5673096945862657+9.
9814174067824841-7.132531566545009+4.3588177707133711-6.3092667884950939.4934416833402861
-3.5712934775899505+-1.539261227422859i-1.204108276406676+-2.868364322298976511-4.861007294
961908+13.336184403162493i-6.684635258816556+-4.9318566861148081i6.6983615319138715.34260
945649307i-6.661104212319744+-6.025474836605843i-13.762193533626547+-13.01929450102414512
5328784947848346+2.0708383919697795i-9.31293965605973+-0.22360425610786194i-16.07493419008
6548+-8.10199310014091i-2.3658401698986147+-4.4769632580336531i0.99807920548292021+-4.253536
258707953i4.545917358271472+-14.5476558883607651-12.729534086694455+2.62637113274645631-2.
8673833478776016+-6.657856329905272i-7.7429818680410021+10.095797518824661-19.7166723904442
7+-12.6344448508570221-5.64099101676735+-3.2547890798173142i-6.9709029622188741+-18.9216272
0033417i-18.588672916793051+9.952739378666884i12.2453039534753181+14.5525952601515561i0.08436
691165709842+-5.806685858236512.4849805290663487+0.15629685730709353i2.61919186122869+-8.9
664697098227551-17.628068995351143+-12.132655820335711-21.5958827109568931-5.0277646576208
921-1.0061810134614624+17.5155107236736312.298314549566676+-8.2617364310312671i0.480603148
2929285+-11.7955377284445361-14.47437643214093+-8.5703058427532021-6.4342536653913841-14.2
9230334800231i-12.835266194584165+0.0015641848492187194i-1.1728514399235266+4.395360945622

CREATING THE UNIQUE REPRESENTATION

- **After obtaining the frequency distribution of the melody (with FFT),**
  - Choose frequency intervals. (0-40, 40-80, 80-120, 120-180, 180-UPPER_LIMIT)
  - For each frequency interval, find the biggest magnitude.
  - In total we'll have 5 numbers which are the peak frequencies for a certain frequency interval.
  - This will be used as the unique representation of the melody.

32	41	106	161	191
33	76	95	123	185
40	68	110	134	232
30	62	88	125	194
34	57	83	121	182
34	42	89	123	182
33	56	99	121	195
30	41	84	146	199

CREATING THE UNIQUE REPRESENTATION

- **Put the song name, region and the composer in hashTable**
- **The hash key (example peak magnitudes: 32 41 106 161 191)**
  - 3141106161191

In short, so far what I have done is to convert the audio file into a 10 digit number.

STEPS TO FOLLOW..

The future work is as follows;

- **Index all the songs in the database and obtain that 10 digit unique (hopefully) key.**
- **Error tolerance helps to increase the match results**
  - Accept any number as same within a certain range.
  - Ex: 17 = 19 (since they are both inside the 14-20 interval)
- **Adding learning algorithms for the songs which are novel**
- **Deciding on weighting system**
- **Deciding on a similarity calculation method.**

CONCLUSION

- **Creating the map of melodies require;**
  - Divide songs into chunks, apply FFT, obtain frequency domain representation.
  - For a couple of frequency intervals, find the peak frequency value.
  - Use those values as the hash key.
  - Train the data
  - Detect novel data

QUESTIONS?