AMATH 482 Project 4 : Music Classification

Yixuan Liu

March 2020

1 Abstract

In daily life, when you heard a clip of music, you can instantly identity its' genre. It is kind of mysterious that how your brain can react to the new information. One way the brain may do is try to match up this clip with some frequencies in your memory to classify the genre of music. In this project, we try to mimic this process of identifying the genre of music by computer.

2 Introduction and Overview

In this project, we mainly use one technique from machine learning: classification. The main goal is to construct a lower-rank subspace and find some threshold on that subspace to distinguish each type. Here we try to train our model to identify the genre of clips of music. We are given three different data set:
First, we use Eminem's rap songs, Josh turner's country songs, and Eagle's rock song to build our model to identify new clips of songs from these bands when these bands are in a different genre.
Second, we use Eminem's rap songs, Dr.Dre's songs, and 2Pac's songs to build our model to identify new clips from these bands when these bands are in the same genre.
Last, we use a variety of songs from different bands within three genres: Rock, Rap, and Country, to train our model to classify these three genres.

3 Theoretical Background

3.1 Singular value decomposition

The image of an n-dimensional unit-sphere can be transformed by a matrix $A \in R^{mxn}$, where m $\geq$ n, to an m-dimensional hyper-ellipse, which has principal semiaxis with length $\sigma_1, \sigma_2, ...\sigma_n$, pointing in direction $\overrightarrow{u_1}, \overrightarrow{u_2}, ...\overrightarrow{u_n}$ for $\overrightarrow{v_1}, \overrightarrow{v_2}, ...\overrightarrow{v_n}$ such that

$$A\overrightarrow{V} = \hat{U}\hat{\Sigma} \tag{1}$$

Where $\overrightarrow{V}$ is a matrix where columns are orthonormal basis. Then, we have

$$A = \hat{U}\hat{\Sigma}\vec{V}^{-1} \tag{2}$$

One useful aspect of svd is that we can reduce our dimensionality of data and reconstruct our data onto the new subspace. Basically, we look at principal components in the matrix V' and choose some of them as our new basis.

3.2 Linear Discriminant Analysis

Let's first define some terminology:

$$S_W = \sum_{i=1}^{n} \sum_{x} (\overrightarrow{x} - \overrightarrow{\mu_i}) * (\overrightarrow{x} - \overrightarrow{\mu_i})^T \tag{3}$$

$$S_B = \sum_{i=1}^{n} m_i (\overrightarrow{\mu_i} - \overrightarrow{\mu}) * (\overrightarrow{\mu_i} - \overrightarrow{\mu})^T \tag{4}$$

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & ... & x_n \end{pmatrix} \tag{5}$$

Where x_i is a column vector representing each measurement of one sample. $m_[i]$ is the number of samples in x_i.$\overrightarrow{\mu_i}$ is the mean of rows of each category(x1 to xm is a category, xm to xk is a category, xk to xn is a category).$\overrightarrow{\mu}$ is the mean of rows in X(total mean). The S_W is a measurement of variance in each category and the S_B is a measurement of co variance of each pair of categories. To best distinguish the difference of these categories, we want to minimize the spread of data in each category and maximize the distance between each group:

$$\overrightarrow{w} = argmax \frac{\overrightarrow{w}^T S_B \overrightarrow{w}}{\overrightarrow{w}^T S_W \overrightarrow{w}} \tag{6}$$

It is equivalent to:

$$S_B \overrightarrow{w} = \lambda S_W \overrightarrow{w} \tag{7}$$

In 3 categories, we get two nonzero eigenvalues which means we should project our data onto two dimensional subspace, which is the first two eigenvectors. Once we have our data projection onto this subspace, we can look at the diagram to find some threshold to distinguish these three categories.

4 Algorithm Implementation and Development

4.1 Loading data

The first step is to create a data set of spectrograms of clips of songs. I choose six songs from each band. Then, I do audioread on each song to get [song, Fs] vector. Since our songs are in stereo and we only need one column for analysis, we drop the right channel. Also,

we resample our song into half the size of origin by resample command. To do so, we can decrease our processing time drastically while keeping data relatively complete. After that, we cut our songs to create more samples: we choose some range of music and cut it into several 5 seconds clips. Then I make the spectrogram of each clip because the spectrogram is a representation of typical frequency, which means it is easier to distinguish some difference from the spectrogram than from the data itself. I resample the spectrogram matrix to a column vector, which corresponding to our x1,x2,...xn in theoretical background.
For the last two task, we basically do the same procedure except choosing three different band's songs within one genre and choosing different songs from three genres from different bands.
The full list of songs is in Appendix C.

4.2 Singular value decomposition

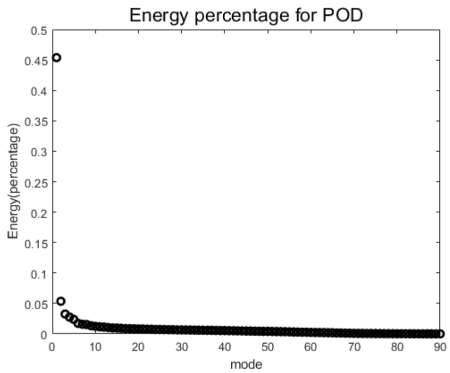
From the loading data part, we get three matrices corresponding to the three categories. Then, I do the SVD on a more significant matrix that contains these three matrices. Then, we plot the energy diagram (Figure 1) to see how many features we need. I choose the first ten features and project our spectrogram onto the primary ten typical principal components. We run the process on the other two tasks and choose different principal components based on figure 1 and some trials.

4.3 Linear discriminant analysis

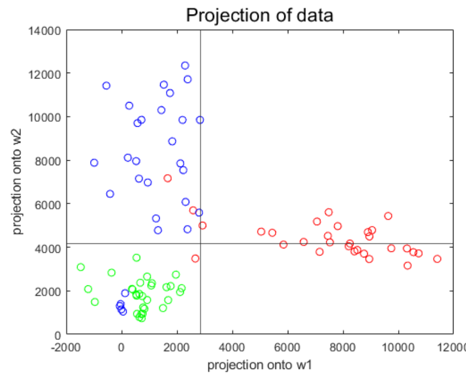
After we have projection data, we can do our Linear discriminant analysis. First we find mean of rows in each projection data matrix. Then we construct SW to save all the variance of each column within each matrix. Then we find all the covariance of two categories' mean, which is saved as SB. Then, we do an eig command to find the eigenvalue of w (see theoretical background for formula). We choose the first two eigenvector as new basis, which called w, because two of eigenvalue are not zero.
Then, we do the same procedure for other two tasks.

4.4 Test data

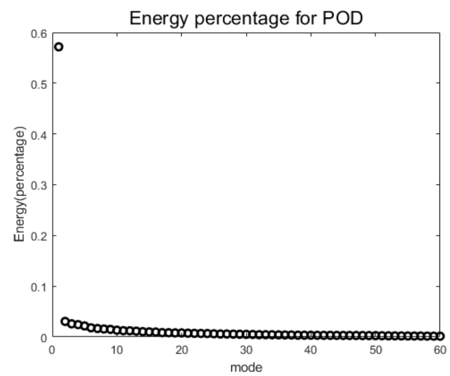
We first set the threshold by looking at the diagram of projection data onto the w. Then, we manually choose the edge of each cluster and set the threshold to be the average of the two edges(see figure 2.4.6). After that, we do the same process to load data with different songs, then project those data onto the principal component space. Finally, we project these data onto the eigenspace and classify theses value with the threshold. In some cases, we compare two components of projection data with two thresholds, and sometimes we just compare one component with one threshold. All the decisions are made manually. There could be an improvement of accuracy if we choose two lines to separate the clusters.



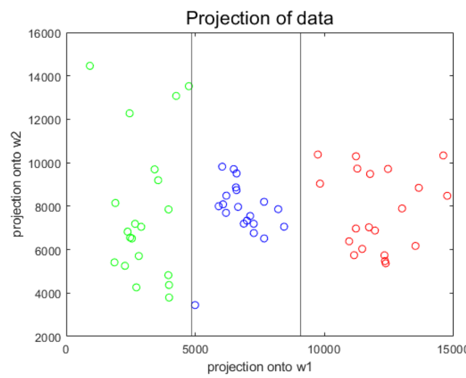
(a) Figure 1 : Energy level for task 1



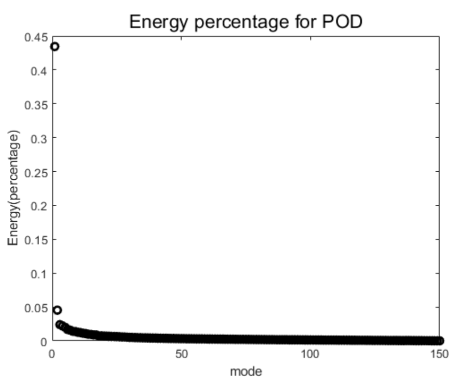
(b) Figure 2 : projection of data



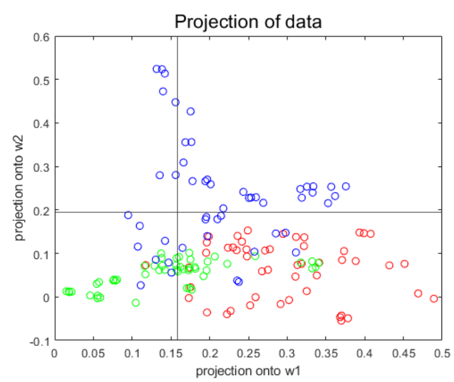
(a) Figure 3 : Energy level for task2



(b) Figure 4 : projection of data



(a) Figure 1 : Energy level for task3



(b) Figure 2 : projection of data

5 Computational Results and Conclusion

For these three tasks, we get 83.3 percentage accuracy for task1, 51.39 percentage accuracy for task2 and 77.33 percentage for task3. Comparing with the first two value, we see a huge decrements when we switch from bands of different genres to different bands with the same genre. What we can see is that the primary feature that distinguishes the difference of songs is probably depend more on genres (style) rather than bands (voice difference). Also, we can tell this fact from third task, which gives us a decent accuracy when we distinguish music based on their genre. We can also increase the accuracy by precisely choosing the threshold, which we do not do well on. We can see that linear discriminant analysis is a power tool to do classification. However, it is still not perfect to make computer act like human. In the future, we expect more and more sophisticated technologies which can benefit the classification and machine learning area.

6 Appendix A: MATLAB functions used and brief implementation explanation

- A = spectrogram(x): return a matrix that represent the spectrogram of x
- [song, Fs] = audioread(strcat(path,'/',filename)): read the music and return the song and the number of sample per seconds
- SVD(X):decompose X in U, V and Σ
- resample(song,Fs/2,Fs): resample the song into Fs/2 samples per seconds.

7 Appendix B: MATLAB codes

```
clc;close all;clear all
%% test 1
e = spec_matrix('./test1',(3:8),(110:5:130));
g = spec_matrix('./test1',(9:14),(110:5:130));
j = spec_matrix('./test1',(15:20),(110:5:130));
[U,S,V] = svd([e g j],'econ');
lambda=diag(S).^2;
figure()
plot(lambda/sum(lambda),'ko','Linewidth',2);
ylabel('Energy(percentage)')
xlabel('mode')
title('Energy percentage for POD','Fontsize',16)
feature = (1:10);%look at the energy diagram
U = U(:,feature);
ne=size(e,2);
ng=size(g,2);
nj=size(j,2);
music = S*V';
es = music(feature,1:ne);
gs = music(feature,ne+1:ng+ne);
js = music(feature,ng+ne+1:ng+ne+nj);
me = mean(es,2);
mg = mean(gs,2);
mj = mean(js,2);
SW = 0;
```

```
for k = 1:ne
    SW = SW+(es(:,k)-me)*(es(:,k)-me)';
end
for k = 1:ng
    SW = SW+(gs(:,k)-mg)*(gs(:,k)-mg)';
end
for k = 1:nj
    SW = SW+(js(:,k)-mj)*(js(:,k)-mj)';
end
mut = mean([es gs js],2);
for i = 1:3
    SB = size(es,2)*(me-mut)*(me-mut)' + size(gs,2)*(mg-mut)*(mg-mut)' +
        size(js,2)*(mj-mut)*(mj-mut)';
end
[V2,D] = eig(SB,SW);
w=V2(:,1:2);%be careful
ve=w'*es;
vg=w'*gs;
vj=w'*js;
figure()
plot(ve(1,:),ve(2,:),'ro'); hold on;
plot(vg(1,:),vg(2,:),'go');
plot(vj(1,:),vj(2,:),'bo');
ylabel('projection onto w2')
xlabel('projection onto w1')
title('Projection of data','Fontsize',16)
thersholdx = (2803+2925)/2;%look at the graph
thersholdy = (3520+4782)/2;%look at the graph
xline(thersholdx)
yline(thersholdy)

teste = U'*spec_matrix('./test1t',(3:6),(110:5:130));
testg = U'*spec_matrix('./test1t',(7:10),(110:5:130));
testj = U'*spec_matrix('./test1t',(11:14),(110:5:130));
teste=w'*teste;
testg=w'*testg;
testj=w'*testj;

ce=0;
cg=0;
cj=0;
for i = 1 :size(teste,2)
    if(teste(1,i)>thersholdx)
       ce = ce+1;
    end
    if(testg(2,i)<thersholdy)
        cg=cg+1;
```

```
    end
    if(testj(2,i)>thersholdy)
        cj=cj+1;
    end
end
acur = (ce+cg+cj)/(3*size(teste,2));

%% test 2
clc;close all;clear all
e = spec_matrix('./test2',(3:6),(30:5:50));
g = spec_matrix('./test2',(7:10),(30:5:50));
j = spec_matrix('./test2',(11:14),(30:5:50));
[U,S,V] = svd([e g j],'econ');
lambda=diag(S).^2;
figure()
plot(lambda/sum(lambda),'ko','Linewidth',2);
ylabel('Energy(percentage)')
xlabel('mode')
title('Energy percentage for POD','Fontsize',16)
feature = (1:10);%look at the energy diagram
U = U(:,feature);
ne=size(e,2);
ng=size(g,2);
nj=size(j,2);
music = S*V';
es = music(feature,1:ne);
gs = music(feature,ne+1:ng+ne);
js = music(feature,ng+ne+1:ng+ne+nj);
me = mean(es,2);
mg = mean(gs,2);
mj = mean(js,2);
SW = 0;
for k = 1:ne
    SW = SW+(es(:,k)-me)*(es(:,k)-me)';
end
for k = 1:ng
    SW = SW+(gs(:,k)-mg)*(gs(:,k)-mg)';
end
for k = 1:nj
    SW = SW+(js(:,k)-mj)*(js(:,k)-mj)';
end
mut = mean([es gs js],2);
for i = 1:3
    SB = size(es,2)*(me-mut)*(me-mut)' + size(gs,2)*(mg-mut)*(mg-mut)' +
        size(js,2)*(mj-mut)*(mj-mut)';
end
[V2,D] = eig(SB,SW);
```

```
w=V2(:,1:2);
ve=w'*es;
vg=w'*gs;
vj=w'*js;
figure()
plot(ve(1,:),ve(2,:),'ro'); hold on;
plot(vg(1,:),vg(2,:),'go');
plot(vj(1,:),vj(2,:),'bo');
ylabel('projection onto w2')
xlabel('projection onto w1')
title('Projection of data','Fontsize',16)
thersholdxr = (8433+9741)/2;%red
thersholdxb = (4750+4988)/2;%blue
thersholdy =(5704+6513)/2;
xline(thersholdxr)
xline(thersholdxb)
teste = U'*spec_matrix('./test2t',(3:6),(20:5:45));
testg = U'*spec_matrix('./test2t',(7:10),(40:5:65));
testj = U'*spec_matrix('./test2t',(11:14),(30:5:55));
teste=w'*teste;
testg=w'*testg;
testj=w'*testj;
% plot(teste(1,:),teste(2,:),'ro')
% plot(testg(1,:),testg(2,:),'go')
% plot(testj(1,:),testj(2,:),'bo')
% ylabel('y-2')
% xlabel('x-1')
ce=0;
cg=0;
cj=0;
for i = 1 :size(teste,2)
    if(teste(1,i)>thersholdxr)
       ce = ce+1;
    end
    if(testg(1,i)<thersholdxb||testg(2,i)<thersholdy)
        cg=cg+1;
    end
    if(thersholdxb<testj(1,i)&& testj(1,i)<thersholdxr)
        cj=cj+1;
    end
end
acur = (ce+cg+cj)/(3*size(teste,2));

%% test 3
clc;close all;clear all
e = spec_matrix('./test3',(3:12),(30:5:50));
g = spec_matrix('./test3',(13:22),(30:5:50));
```

```
j = spec_matrix('./test3',(23:32),(30:5:50));
[U,S,V] = svd([e g j],'econ');
lambda=diag(S).^2;
figure()
plot(lambda/sum(lambda),'ko','Linewidth',2);
ylabel('Energy(percentage)')
xlabel('mode')
title('Energy percentage for POD','Fontsize',16)
feature = (1:2);%look at the energy diagram
U = U(:,feature);
ne=size(e,2);
ng=size(g,2);
nj=size(j,2);
music = S*V';
es = music(feature,1:ne);
gs = music(feature,ne+1:ng+ne);
js = music(feature,ng+ne+1:ng+ne+nj);
me = mean(es,2);
mg = mean(gs,2);
mj = mean(js,2);
SW = 0;
for k = 1:ne
    SW = SW+(es(:,k)-me)*(es(:,k)-me)';
end
for k = 1:ng
    SW = SW+(gs(:,k)-mg)*(gs(:,k)-mg)';
end
for k = 1:nj
    SW = SW+(js(:,k)-mj)*(js(:,k)-mj)';
end
mut = mean([es gs js],2);
for i = 1:3
    SB = size(es,2)*(me-mut)*(me-mut)' + size(gs,2)*(mg-mut)*(mg-mut)' +
        size(js,2)*(mj-mut)*(mj-mut)';
end
[V2,D] = eig(SB,SW);
w=V2(:,1:2);
ve=w'*es;
vg=w'*gs;
vj=w'*js;
figure()
plot(ve(1,:),ve(2,:),'ro'); hold on;
plot(vg(1,:),vg(2,:),'go');
plot(vj(1,:),vj(2,:),'bo');
ylabel('projection onto w2')
xlabel('projection onto w1')
title('Projection of data','Fontsize',16)
```

```
thersholdxb = (0.1921+0.1954)/2;%blue
thersholdy =(0.1635+0.1528)/2;
yline(thersholdxb)
xline(thersholdy)

teste = U'*spec_matrix('./test3',(3:12),(70:5:90));
testg = U'*spec_matrix('./test3',(13:22),(70:5:90));
testj = U'*spec_matrix('./test3',(23:32),(70:5:90));
teste=w'*teste;
testg=w'*testg;
testj=w'*testj;
% plot(teste(1,:),teste(2,:),'ro')
% plot(testg(1,:),testg(2,:),'go')
% plot(testj(1,:),testj(2,:),'bo')
% ylabel('y-2')
% xlabel('x-1')
ce=0;
cg=0;
cj=0;
for i = 1 :size(teste,2)
    if(teste(1,i)>thersholdxb && teste(2,i)<thersholdy)
       ce = ce+1;
    end
    if(testg(1,i)<thersholdxb && testg(2,i)<thersholdy)
        cg=cg+1;
    end
    if(testj(2,i)>thersholdy)
        cj=cj+1;
    end
end
acur = (ce+cg+cj)/(3*size(teste,2));

function [A] = spec_matrix(path,numofsong,timeclips)
folder = dir(path);
A=[];
for i = numofsong
    filename=folder(i).name;
    [song, Fs] = audioread(strcat(path,'/',filename));
    song = song(:,1);
    song = resample(song,Fs/2,Fs);
    Fs = Fs/2;
    %p8 = audioplayer(song(Fs*100:Fs*110),Fs);playblocking(p8);
    for j=timeclips
        temp = song(Fs*j:Fs*(j+5));
        temp_sp = abs(spectrogram(temp));
        temp_sp = reshape(temp_sp,length(temp_sp(:)),1);
        A = [A temp_sp];
```

```
    end
end
end
```

7.1 Appendix C: List of songs

For task 1 training set:

- Premonition-Eminem from 110s to 130s-Rap
- Lucky you-Eminem from 110s to 130s-Rap
- Godzilla-Eminem from 110s to 130s-Rap
- Space bound-Eminem from 110s to 130s-Rap
- Lose yourself-Eminem from 110s to 130s-Rap
- Rap god-Eminem from 110s to 130s-Rap
- Desperado-Eagles from 110s to 130s-Rock
- Hotel California-Eagles from 110s to 130s-Rock
- Take it Easy-Eagles from 110s to 130s-Rock
- Life in the fast lane-Eagles from 110s to 130s-Rock
- One of these nights-Eagles from 110s to 130s-Rock
- New kid in Town-Eagles from 110s to 130s-Rock
- Firecracker-josh Turner-from 110s to 130s-country
- Your Man-josh Turner-from 110s to 130s-country
- Another Try-josh Turner-from 110s to 130s-country
- Your Smile-josh Turner-from 110s to 130s-country
- Long Black Train-josh Turner-from 110s to 130s-country
- Why Don't We just Dance-josh Turner-from 110s to 130s-country

For task 1 testing set:

- Homicide-Eminem from 110s to 130s-Rap
- Fall-Eminem from 110s to 130s-Rap
- Normal-Eminem from 110s to 130s-Rap

- Love the way you lie-Eminem from 110s to 130s-Rap
- Wasted time-Eagles from 110s to 130s-Rock
- Victim Of Love-Eagles from 110s to 130s-Rock
- Try and love again-Eagles from 110s to 130s-Rock
- The Last Resort-Eagles from 110s to 130s-Rock
- Would you go with me-josh Turner-from 110s to 130s-country
- Howntown Girl-josh Turner-from 110s to 130s-country
- I wouldn't Be A Man-josh Turner-from 110s to 130s-country
- The Answerr-josh Turner-from 110s to 130s-country

For task 2 training set;

- Forgot about Dre-Dr Dre from 30s to 50s-Rap
- Still Dre-Dr Dre from 30s to 50s-Rap
- Talking To My Diary-Dr Dre from 30s to 50s-Rap
- What's the difference-Dr Dre from 30s to 50s-Rap
- Not Afraid-Eminem from 30s to 50s-Rap
- Space Bound-Eminem from 30s to 50s-Rap
- Normal-Eminem from 30s to 50s-Rap
- Love the way you lie-Eminem from 30s to 50s-Rap
- Ambitionz Az A Ridah-2Pac from 30s to 50s-Rap
- Do for love-2Pac from 30s to 50s-Rap
- All About U-2Pac from 30s to 50s-Rap
- All Eyez On Me-2Pac from 30s to 50s-Rap

For task 2 testing set;

- Kush-Dr Dre from 20s to 45s-Rap
- The Watcher-Dr Dre from 20s to 45s-Rap
- Bang Bang-Dr Dre from 20s to 45s-Rap
- Light Speed-Dr Dre from 20s to 45s-Rap

- Fall-Eminem from 40s to 65s-Rap
- Venom-Eminem from 40s to 65s-Rap
- The way I am-Eminem from 40s to 65s-Rap
- Beautiful-Eminem from 40s to 65s-Rap
- Changes-2Pac from 30s to 55s-Rap
- Dear Mama-2Pac from 30s to 55s-Rap
- Skandalouz-2Pac from 30s to 55s-Rap
- Life Goes On-2Pac from 30s to 55s-Rap

For task 3 training set:

- Lonely if you are-Chase Rice from 30s to 50s-country
- Cool Anymore-Jordan Davis from 30s to 50s-country
- homecoming queen-Kelsea Ballerini from 30s to 50s-country
- kinfolks-Sam Hunt from 30s to 50s-country
- Getting Good-Lauren Alaina from 30s to 50s-country
- Nothing Like you-Luke Combs from 30s to 50s-country
- Chasin' You-Morgan Wallen from 30s to 50s-country
- Like I knew You Would-Payton Smith from 30s to 50s-country
- Catch-Brett Young from 30s to 50s-country
- The Bones-Maren Morris from 30s to 50s-country
- We will Rock You-Queen from 30s to 50s-Rock
- t.n.t-ac/dc from 30s to 50s-Rock
- Welcome to the jungle-Guns N'Roses from 30s to 50s-Rock
- Crazy Train-Ozzy Osbourne from 30s to 50s-Rock
- Iron Man-Black Sabbath from 30s to 50s-Rock
- Desperado-Eagles from 30s to 50s-Rock
- Hotel California-Eagles from 30s to 50s-Rock
- Back in the Saddle Again-Aerosmith from 30s to 50s-Rock

- Born in the USA-Bruce Springsteen from 30s to 50s-Rock
- Who are you-The Who from 30s to 50s-Rock
- Forgot about Dre-Dr Dre from 30s to 50s-Rap
- Do for love-2Pac from 30s to 50s-Rap
- Ambitionz Az A Ridah-2Pac from 30s to 50s-Rap
- Space Bound-Eminem from 30s to 50s-Rap
- Not Afraid-Eminem from 30s to 50s-Rap
- Still Dre-Dr Dre from 30s to 50s-Rap
- Sum 2 Prove-Lil Baby from 30s to 50s-Rap
- Woah 2 Prove-Lil Baby from 30s to 50s-Rap
- Baby Birkin-Gunna from 30s to 50s-Rap
- drip too hard-Gunna from 30s to 50s-Rap

For task 3 testing set:

- Lonely if you are-Chase Rice from 70s to 90s-country
- Cool Anymore-Jordan Davis from 70s to 90s-country
- homecoming queen-Kelsea Ballerini from 70s to 90s-country
- kinfolks-Sam Hunt from 70s to 90s-country
- Getting Good-Lauren Alaina from 70s to 90s-country
- Nothing Like you-Luke Combs from 70s to 90s-country
- Chasin' You-Morgan Wallen from 70s to 90s-country
- Like I knew You Would-Payton Smith from 70s to 90s-country
- Catch-Brett Young from 70s to 90s-country
- The Bones-Maren Morris from 70s to 90s-country
- We will Rock You-Queen from 70s to 90s-Rock
- t.n.t-ac/dc from 70s to 90s-Rock
- Welcome to the jungle-Guns N'Roses from 70s to 90s-Rock
- Crazy Train-Ozzy Osbourne from 70s to 90s-Rock

- Iron Man-Black Sabbath from 70s to 90s-Rock
- Desperado-Eagles from 70s to 90s-Rock
- Hotel California-Eagles from 70s to 90s-Rock
- Back in the Saddle Again-Aerosmith from 70s to 90s-Rock
- Born in the USA-Bruce Springsteen from 70s to 90s-Rock
- Who are you-The Who from 70s to 90s-Rock
- Forgot about Dre-Dr Dre from 70s to 90s-Rap
- Do for love-2Pac from 70s to 90s-Rap
- Ambitionz Az A Ridah-2Pac from 70s to 90s-Rap
- Space Bound-Eminem from 70s to 90ss-Rap
- Not Afraid-Eminem from 70s to 90s-Rap
- Still Dre-Dr Dre from 70s to 90s-Rap
- Sum 2 Prove-Lil Baby from 70s to 90s-Rap
- Woah 2 Prove-Lil Baby from 70s to 90s-Rap
- Baby Birkin-Gunna from 70s to 90s-Rap
- drip too hard-Gunna from 70s to 90s-Rap