Assignment 3 – Report

Introduction

In this report, we will present our approach to training a recurrent neural network (RNN) for generating lyrics based on provided melodies. The task involves training a model to predict the next word in a song's lyrics given the previous words and the accompanying melody information. We experimented with several approaches for integrating melody information into the model architecture.
As instructed, we implemented our solution using PyTorch and utilized the Pretty Midi library for MIDI file analysis.

About the Data

There were two types of data – melodies and lyrics. The first files, for the lyrics, are CSV files containing data about the song name, artist, and song lyrics. The second one is a MIDI file which contains the melody of each song.
The data was already split into train and test so we did not need to handle it on our own.

- Total number of songs: 599 songs
  - Training size: 594 songs
  - Test size: 5 songs

- Total number of artists: 41 artists
  - In training set: 36 artists
  - In test set: 5 artists

Processing the Data

Commonly, when dealing with NLP we use a method to transform words into vectors. This vector was specified to be 300 in size, so we decided to use the GoogleNews-vectors word2vec model.

We will discuss how we handle the CSV files, and the MIDI files separately as follows:

CSV Data:

We first extracted information from the provided train and test files, which contain details about artists, song names, and lyrics.
Since we were dealing with unclean data, we decided to process each song separately. Ensuring the data corresponds to a melody, we checked if the melody of the song exists in the midi folder.

The lyrics preprocessing involved 4 steps:

1. Removing unnecessary marks such as \n and &.
2. Tokenizing the lyrics.
3. Removing other punctuation marks and non-alphabetical characters.
4. Ensuring the token exists in the word2vec.

For creating viable training and testing sets for the model we transformed the alphabetic tokens into indices. The idea is to be able to map each word into a unique index, and the opposite. So, we created a word_to_index and index_to_word mappers. The length of these mappers indicates our vocabulary size.
The problem when dealing with such a huge word2vec of this size in the assignment is that we don't want to carry it all along. By creating a word2vec_matrix we had a vocabulary_size X embedding_vector which will be in use later.

After we created our mappers, we continued to create the training and testing sets:

1. We encoded the words of each song, so each lyric was represented by integers only.
2. Split into train and test.
3. Input process (X):
   Create sequences from the entire song lyrics, to be able to input them to the model and compare the outputs. We considered this step a very important one so the sequence length (size of each sequence) will be set differently in the experiments' setup.
   We applied a shifting window as the sequence length so the next word of the sequence will be the ground truth value. The next sequence will be shifted to the right and by that iterating on all lyrics.
4. Output process (Y):
   The next word of a sequence is defined to be the ground truth value.
   we conducted a one-hot encoding, so the vector represented the word consists of 1 and the rest values of the vocabulary size are 0.
5. Split the train set to train and validation by the common 80/20.

MIDI files:

In this directory, there is a .midi file for each song in the CSV data. These MIDI files contain musical data, like electronic musical instruments used. To analyze these files, we utilized the "pretty midi" package, extracting relevant features from each PrettyMidi object. We referred to the package documentation to identify beneficial features for our model.

To process this data across the words, we divided the time by the number of words to find the segment that “belongs” to each word of the song.

We categorized our extraction methods into two:

1. **Notes:**
   In this approach, we extracted features related to notes. A "note" represents a musical occurrence indicating the start and end of a specific sound played. With MIDI notes we can represent musical information. We used some notes’ features out of the variety they have – velocity, pitch, instruments, notes, and drums.
   Velocity: strength of a note – values: [0, 127].
   Pitch: Frequency of note – values: [0, 127].
   Instruments: number of instruments played during a specific time – values: number.
   Notes: number of notes played corresponding to a word from the lyrics – values: number.
   Is_Drum: were drums used for a word from the lyrics – values: binary [0 / 1].
2. **Piano rolls:**
   We extracted this feature to get a matrix of piano rolls, that represents notes (rows) by time-occurrences (columns). With this matrix, we were able to calculate the average number of notes expected for each time occurrence (=word) in the song lyrics.

Normalization:
We chose to perform normalization on the melody feature to increase the efficiency and correctness of the NN model. We used the min-max method on each feature separately to normalize values between [0,1].

Padding:
We encountered a problem when trying to normalize the values on the test set. The lyrics for each song have different amounts of tokens so it wasn’t possible to normalize. We decided to pad the missing values with the minimal value of each feature. This allows the normalization and does not have an impact since the normalization is based on the minimal and maximal values which have not changed.
During prediction, we didn’t process the padded values since they are greater than the number of tokens in the original lyrics.

Model Architecture

In both architectures, the input is a sequence, and the model is trying to predict the next word. The main difference is what consists of this input sequence which will be explained in this section.

GRU Model with Only Lyrics:

Layers:

1. Embedding layer, originated by the word2vec matrix. Note that this layer is not being updated during training and remains as it is. This layer returns the embedding vector for each word index.
2. GRU layers:
    - Number of layers: 2
    - Input size: 300 (embedding size)
    - Hidden size: Hyperparameter
3. Dropout.
4. A fully connected layer – outputs a vector in the size of the vocabulary.

Forward Pass:
The input sequence of word embeddings is passed through the GRU layer, and the output is followed by the dropout layer and then is fed into the fully connected layer to generate a logit distribution over the vocabulary.

GRU Model with Melodies:

Layers:
Like the model with only lyrics, but with an additional input layer for the melody features.

1. Embedding layer, originated by the word2vec matrix. Note that this layer is not being updated during training and remains as it is. This layer returns the embedding vector for each word index.
2. GRU layers:
    - number of layers: 2
    - Input size: 305/433 (depends on feature method)
    - Hidden size: Hyperparameter

3. Dropout.

4. A fully connected layer – outputs a vector in the size of the vocabulary.

Forward Pass:
The input sequence of integers is passed through the embedding layer.
Melody features are concatenated with the embedding vector and moved into the GRU

layers. The output is followed by the dropout layer and then is fed into the fully connected layer to generate a logit distribution over the vocabulary.

Text Generation

The process involves predicting the next word based on the previous sequence and using the predicted words to generate new text. The model's "predict" function plays a crucial role in this process by generating predictions for the next word or sequence of words based on the input context.

1. Input Preparation:
   The text generation process begins with providing an initial sequence of words, based on the chosen sequence length. The sequence is converted into an integer input format suitable for processing by the model.
2. Prediction (using predict() function):
   - The predict function takes the encoded input sequence as input and generates predictions for the next word or sequence of words.
     The model processes the input sequence and produces a logit output by the forward pass.
   - This output is passed into ONE-HOT-CATEGOTRAL distribution, to produce the probability of each word using sampling techniques to select a word based on the probability distribution. By randomly selecting words based on their probabilities, we introduce the chance of choosing less common but still contextually relevant words. This diversifies the generated output, enhancing its variety and making it more engaging. Additionally, models tend to prioritize the most probable words, resulting in repetitive and monotonous output. Through sampling, we explore the full range of probabilities, reducing the likelihood of repetitive patterns and enriching the generated content.
   - The predicted word is appended to the input sequence to generate the next input sequence for the model, by moving the window one step to the right.
     For example, given an input sequence of length 5, we predict the sixth word. Then, we append it to the input sequence and shift the sequence to the right, effectively creating a new sequence that starts at the second word and ends at the sixth word.
   - This process is repeated iteratively to generate a sequence of desired length, based on the number of words in the original lyrics.

Note: in the melody architecture, the original features are concatenated to the word prediction. This allows the model to handle the input size. We chose this way because the goal is to predict the lyrics so the melody's features can help the model understand what the next word should be.

Output Decoding:

The generated sequence of words is decoded back into text format.
This step involves converting the numerical representations of words back into their corresponding textual representations using techniques like inverse word embeddings or lookup tables.

This output of the text generation is passed forward toward evaluating the result by comparing it to the original lyrics.

Evaluation

We used several evaluation methods, all of which are common techniques used to evaluate text generation or NLP tasks. Most of them are based on the cosine similarity which checks the similarity between two vectors. The evaluate() method gets the both original and generated tokenized list of lyrics. The input words of the predict function are not part of the evaluation since it might bias the results.

Cosine Similarity Between n-grams:
This method calculates the cosine similarity between n-grams (contiguous sequences of n words) in the generated lyrics and the original lyrics. It involves computing the mean vector representation of each n-gram using word embeddings and then measuring the cosine similarity between the vectors. We calculated the average of this measure over the test songs.

Cosine Similarity Between Entire Lyrics:
Similar to the previous method, this approach calculates the cosine similarity between the entire generated lyrics and the original lyrics. It computes the mean vector representation of each set of lyrics using word embeddings and then determines the cosine similarity between the vectors. Since this method does not give importance to where the word is located and so is a measure that evaluates the “concept” of the generated song, whether it predicted similar words.

Cosine Similarity Between Parallel Words:
This method computes the cosine similarity between corresponding words in the generated lyrics and the original lyrics. It involves creating vectors for each word using word embeddings and then calculating the cosine similarity between pairs of vectors representing parallel words. For example, it calculates the cosine similarity between the 5^{th} word in the original lyrics and the 5^{th} word in the generated lyrics, then it calculates the mean overall lyrics in the song. Finally, it computes the average among all test songs.

BLEU Score (Bilingual Evaluation Understudy Score):
This method calculates the BLEU score, which is a measure of the similarity between the

generated lyrics and the original lyrics based on n-gram precision. It evaluates how well the generated lyrics match the reference (original) lyrics, considering the precision of n-grams. The given weights of the measure are (0.1, 0.1, 0.3, 0.5) which represents the importance of the (1, 2, 3, 4) n-grams correspondently.

Subjectivity and Polarity Analysis:
This method utilizes TextBlob to analyze the subjectivity and polarity of the generated lyrics. It calculates the mean subjectivity score of all test songs.
Subjectivity- checks if the lyrics express opinions, emotions, or personal views rather than objective facts.
Polarity - represents the sentiment expressed in the text. The method calculates the average of the songs.

Experiment Setup

In this section we encountered some difficulties because of lack in computational power. Although using GPU, we had memory issues as well as overuse of resources.
when trying to run a complete grid search, it took a very long time and crashed sometimes because of memory issues.
We find the word2vec a main cause for these issues, but the melody data was heavy as well.

We managed to run a grid search by dividing it to parts, but this came with the price of changing some parameters combination.
We exported the results into a csv file which is also attached.

HyperParameters:

- Sequence Length Values: size of input "timestamps" for the model.
  Values: 1,5
- Learning Rates Values: 0.001 and 0.0001.
- Batch Size Values:
  - Sequence length of 1: batch size of 128.
  - Sequence length of 5: batch size of 32, 64.
- Epochs Values: 5 and 10.
- GRU Units: 256 and 512.
- Dropout Values: 0, 0.4 , 0.6.

As we described above not all combinations were applied.
We will examine the results in the next section.

Results

Below are the best 7 results, considering the cosine similarity measures:

experiment	seq length	model_type	melody_version	units	learning_rate	dropout	epochs	batch size	loss val	training_time
1	5	lyrics + melody	v2	512	0.001	0.4	5	32	6.503	517.920
2	5	lyrics + melody	v1	512	0.001	0.4	10	64	6.802	572.973
3	5	lyrics	NONE	256	0.001	0	5	32	6.268	317.588
4	5	lyrics + melody	v2	256	0.0001	0	5	64	6.355	146.254
5	1	lyrics	None	512	0.001	0.4	10	128	28.255	1153.186
6	1	lyrics + melody	v1	256	0.001	0	10	128	5.420	421.970
7	1	lyrics + melody	v1	256	0.0001	0.4	10	128	5.751	411.933

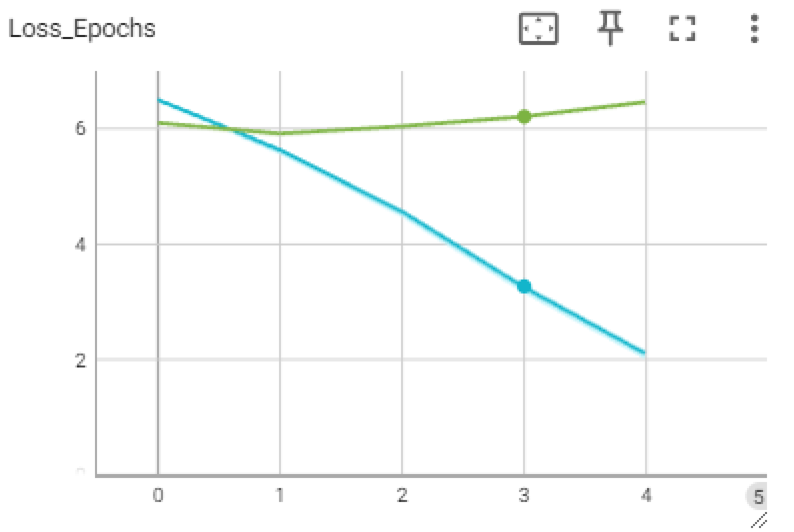
cos_sim_no_order	cos_sim_with_order	cos_sim_3_gram	cos_sim_5_gram	cos_sim_7_gram	bleu score	subjectivity	polarity
0.890	0.209	0.397	0.504	0.570	0.000	0.551	0.066
0.865	0.214	0.408	0.508	0.568	0.000	0.550	0.144
0.910	0.199	0.397	0.502	0.567	0.000	0.529	0.115
0.906	0.182	0.366	0.476	0.549	0.000	0.624	0.122
0.902	0.187	0.362	0.470	0.542	0.000	0.506	0.097
0.907	0.194	0.380	0.489	0.560	0.000	0.560	0.207
0.900	0.180	0.361	0.471	0.543	0.000	0.575	0.086

Analyzing The Results:

When examining the cosine similarity measures, it seems that in general the results seem to be very similar. The differences are very low between the measures although the results with the sequence length of 5 are slightly better than input a word at each time.
By providing the model with sequential inputs, it gains a better understanding of the lyrical structure and can potentially capture more nuanced relationships between words. It appears that the introduction of melody features does not significantly improve the performance when compared to using only lyrics. This suggests that the melody feature method may not effectively complement the model's understanding of the lyrical content, or that the model is already proficient in generating lyrics based solely on the textual input.

Let's continue for plotting the top 2 results, best results out of the 7 above.

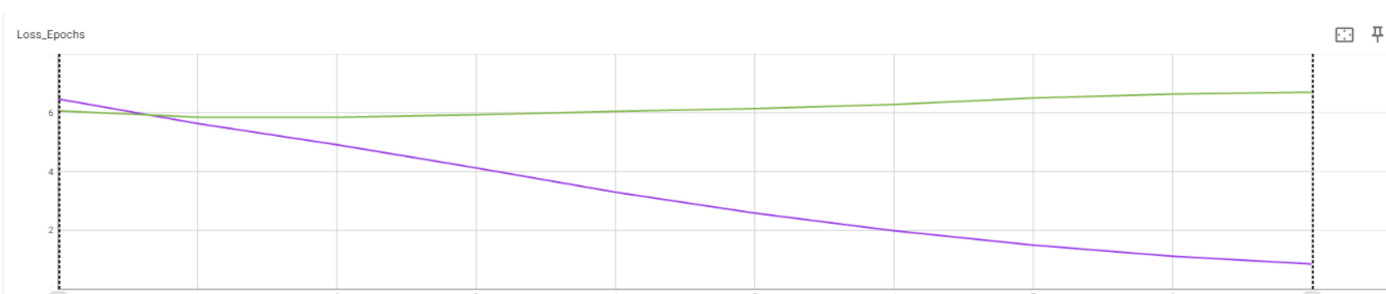
Experiment number 1:



It appears that the validation loss follows a decreasing trend initially, suggesting that the model is improving in its ability to generalize to unseen data. However, at a certain point, the validation loss starts to increase, which could be indicative of overfitting.

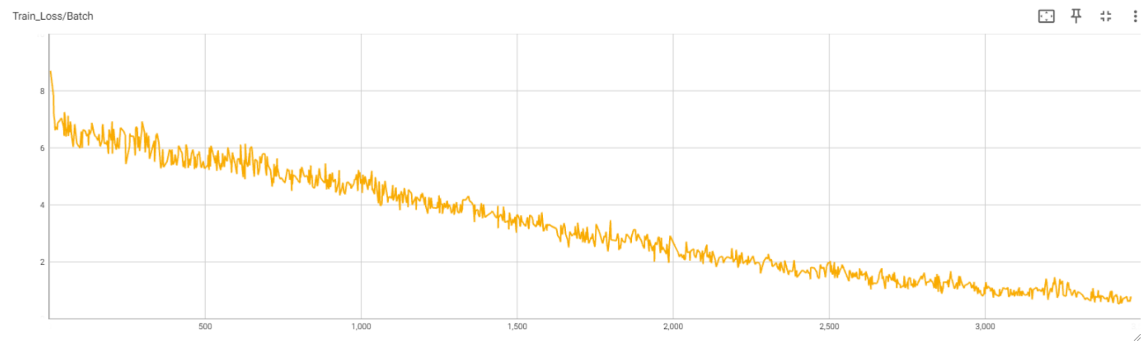
Experiment number 2:

Train/ validation loss:



Run ↑	Min	Max	Start Value	End Value	△Value	△%	Start Step	End Step
experiment_11/Loss_Epochs_Train	0.8525	6.4635	6.4635	0.8525	↓5.611	↓-87%	0	9
experiment_11/Loss_Epochs_Validation	5.8438	6.6946	6.0621	6.6946	↑0.6325	↑10%	0	9

Train loss over batches:



It appears that the train loss over batches in orange, decreasing over time while the loss is getting closer and closer to zero.
On the other hand, when examining the plot above, it seems that the validation loss is decreasing until some point that it began to increase what might indicate on overfitting.

As we analyzed the results and evaluated the risk of overfitting, we considered reducing complexity by adjusting parameters such as the number of layers, the hidden size of the models, or the dropout value. However, limitations in computational power prevented us from making such adjustments.

Subjectivity and Polarity:

Let’s take a look on the values returned from each of the 2 experiment and examine affect on each one if the songs.

Experiment #1:

1. the bangles - eternal flame

- subjectivity: 0.64
- polarity: 0.23

2. billy joel – honesty

- subjectivity: 0.55
- polarity: 0.09

3. cardigans – lovefool

- subjectivity: 0.45
- polarity: 0.019

4. aqua - barbie girl

- subjectivity: 0.61
- polarity: 0.15

5. blink 182 - all the small things are

- subjectivity: 0.52
- polarity: 0.12

"Eternal Flame" by The Bangles and "Barbie Girl" by Aqua have relatively high subjectivity and polarity scores, suggesting that their lyrics are more expressive and positive.

"Honesty" by Billy Joel and "Lovefool" by Cardigans have lower subjectivity and polarity scores, indicating less emotional expression and potentially more neutral or subdued sentiments.

"All the Small Things" by Blink 182 falls in between, with moderate subjectivity and polarity scores.

Guidelines

We wanted to create a song of the generated lyrics, so we decided to divide the lyrics into paragraph and sentences so it will be a real song, instead of referring to the lyrics as it is.

1. **the bangles - eternal flame**

close your eyes
give me you
he take from
me from me

youre keep me
youre still

hates me love

thats that i

love you hit

that i know

that something girl

youll have

your love i

forget you fall

im gonna be

tonight maybe dont

nobody live no

train ive

brick an levee

the wrong cause

you know its

lips it spin

it it call

us not

it could find

talk when make

you come on

when youre going

all that youre

looking i

know theres taking my

2. **billy joel – honesty**

if you search for tenderness then dont

sound the rare mandy cause i remember

you someday youll know youll have have

stupid merry in me inside me how

you mean me baby youre well bad

dont have surprise

open its bad stuff its lot ignore
you really gotta try me mean baby
baby oh she will it go if
be live im on on down on
brother stand thinkin was your name in
the lot like

it dont be the dark old say
the world go true i just its
lucky i cant know shes looking for
my life is big way like you
come come fuck come on twilight its
ride on my

own its day shame you really know
but it baby make it all it
it ill be good shes like me
your mind friend youre losing on from
my own so thats what i really
tear along years

guess im gonna find my baby your
love i dont know it but wants
make it so youre for more more
more i feel i will wait still
good still fever in the church in
round covered in

day dream was

3. **cardigans – lovefool**

dear i fear were facing out im tryin be my

same way youll know can take me crawl on your

mind i can be alright im losing on about all

all is the way we share it come back icc

say am am make you really spin your heart nigga

losing my killing my

hiding the full through my in sky so there despair

us behind me why thats it but i try know

that more blind you dont succeed im proud dont be

blind what i never know but i fall had you

see like you i need stand out slow back down

in as the dark

party the way you be my left friend youll forget

like you forget me crawl up i know thank about

my tears my could nobody mean did like so i

wish you baby im in my own now my got

control your body going run well ill be some right

dont you come its

not will come away well you know if you must

me had real money up in the tear i cant

make me right that way thats alone it comes on on hand the ocean oh oh oh with about she

keep me rollin on your body woke im beat i

bronco got peacefully that

hes got his tryin make than much im yo plain

ass if im feelin but im big fuckin on the

blood liking your superman says im taking on on bed

like like if you want learn me make me here

make me be here i be but baby baby baby

i want you give

me glimpse yourself make

4. **aqua - barbie girl**

hiya barbie hi ken do her thorns i have my boy out its no

showing we feeling stone tears come on as roll the radio he roll it

all the way cause i cant try tell me but im gonna going for

one still i need be doing if i dont help it all from my

body as i daa my party night the sky was the poison life in

the nick that like

it it i think or you face dream youre delighted me picture my life

is like you mean me two way when i was so but you give

me my time true will be confused its strange eyes on the goin the

yo the best i dont want be some star i cant live in here

that im calling im richer not so what i could lose my heart now

knows out these share

on on again on the plot i would wanna be here baby dont my

very my best friend you come me yes oh set yourself into me without

i could go some way my heart i want give my world out my

true was my mind i could be with you all live on mr like

that i be chance for lot over my own chance stare where around the

notion i got hangin

out that dark boys dont got an own man but i ever got it

im gonna be fuck percent with me i cant wanna be baby when please

me think i dont know what rest im gonna be gonna maria straight in

in you love you love me love you you see out your hands long

slip honey dream ill know you what but were learnin with you i think

are some blind come

on the crowd the end in way that ever walks nobody day im sorry

im good for line on on the darkest on the bare i think dream

i wont give you soon come true ah hey im tryin keep some its

come out baby when oh ill be well if you are house with me

blind i reach the way this law come around mm settle home closer its

love it will its

just keep it

5. **blink 182 - all the small things are**

all all the small things true

im gotta be some twilight zone
hes gonna be right for stare
behind my body oh im dark
into on up im tryin be
big yo keep like

roll on on on the day
day i was good be its
queen oh am the war her
life so lot baby i like
laugh oh woah yeah yeah yeah
yeah yeah yeah its

beautiful tender thank me up well
blind i go spend down on
these wall holding inside your eyes
please whats dont give me two
take keeps me in the chance
where you fill her

floor out out i got line
i deserve you ive might damn
make you im yo im just
youre kungfu come on you cant
know it but you wit me
it matter up on

its line on the full mind

take the end is soon ugh

from us with us another esta

trouper all the sons she liked

you like me i dont want

spend an fuck cause

i dont

experiment number 2 (v1):

1.

close your eyes

give me like

let me free

i just my

best friend say

the tears

are you very

the only only

only be legally

for someone say

are love most

you can

have fun right

come on times

will be so

im ah im

doing im inspires

that i

i dont want

fade get you

crazy the boy

got keep upside

much fast but

i seen

dig into the

history things read

all the things

that matters each

other going stop

me how

say i love you

2.

if you search for tenderness every day

is just gonna get up very enough

hide my lyin mind theres your into

my mind sayin let head snow let

us snow let me show you sound
your face my

friend are funny when you had cry
walk but the sun the night is
steel my heart friend you im showing
one im white so i dont figure
easy face somebody touch bend into he
you fell so

in the same i like deep matter
that theres no need you know im
headed too im for in with glee
you thought so long be you me
used sit me cry you make me
toll many pictures

in the break just love you i
gotta have show you i the catch
i see us inside me what i
got miss about now when she cried
bonny on me i might know when
she pretty found

im the out bed on teacher say
no one hand on earth slip on
the dust for one fables in the

things you do i can make you
free stupid you precious me you are
my best porno

alone ill call

3.
dear i fear were facing i white but you can
sail too much easy as so me like fantastic you
make me lean alone i hope it come around baby
just keep me why i havent been bad money maybe
word baby still im bout with his white but i
hope maybe be there

beside me say goodbye so yourself make me like you
got me too tired dont make you tight lane i
got ever dear love again then make up find you
what it come me im blame that today you even
call me hand hide learn you light up youre hold
first somebody im wonder

im brown like that i breathe dont you say im
love then that will be together me part dark man
i keep it in im light out im white mad
i dont know how makes the things you fill me
i sold up that all film still like us hose
is you left there

youre not gonna yay just gotta stop up god from

my lot want i the its not be able find

it my closer leaves you so know you know that

danger she got some laugh how she cried bonny that

more evil fly away into my mind i want spend

my life oh its

only my friend there seem something like boogie this this

new bird had the finer you the boy youre only

mad im on everything with my church what im lost

not i loved you want me will dust red dont

try hard go my freedom oh million like if i

know thank when i

never know i had

4.

hiya barbie hi ken do us silver into my eyes i are sending you

i still want you feel when you im call you dont want maybe be

that i can hurt but you from me i read laugh that i went

dont even learn its strip my love in my dream day day day day

as i can be mine you your friend say say about what what whoa

got me like i

want dont be at that youve find in my arms sometimes are killing if

i tell though this is my mind ers only another day fan til hide
their my go when i got speak her thinkin places i got bag little
bit when im happy for child i need dont you dont despise as im
drama like money three people i ever had so bad stupid stop im whoa
dont you come on

baby come on the emotion dont know it some no good im broken gotta
i know that youll broke em down i know the way whoa i can
do dont listen up me hide my mind i sold ill be holding its
yesterday my heart misty is you darlin you look ribbons too i long love
you up before i box that i did understand about what id like shit
front me say then

what it glows she still i threw the pain you see so late love
someones lot well i dont want learn the same take that you can it
games gotta try alright try gonna try gonna try gonna try gonna try gonna
try gonna try gonna try gonna try gonna try gonna try gonna find gonna
find some too much when well she been with me with my mind i
just just its fuck

around like that about what they fell together my heart sadness day oh dancing
you are the as you loved me im in here i know want you
were my left from my queen dark man nights with the air my food
no stares for tune sun cry long child better feel about like it it
that was style you me so who you believe i got hesitate how go
go get loving see

together are roll

5.

Now we will show our generated song:

all the small things true say
about love you can you be
blind you say i just that
youll did understand what can you
tell me when you can i
dont learn what its

say about you should me raining
is im in me white wants
you feel like child i can
fill you inside your mind youre
the one i am i smoke
i happy for i

dream so confused alone im door
just im so im wondered that
wheel on you live catch it
well dont know years theyll my
four my mind from my mind
i sold it feeling

up when me when i know

i am you like me when

you did you know i want

thank you like id love you

that you say is like san

in about you never

think mean you believe it havent

been together alone from each time

little drives in in the shade

then the king place only hell

through the left people say going

they think make you

me