Playlist Generation using Emotion in Music

Christine Chao, Sophia Mallaro, Ming-Chang Chiu, Sabyasachee Baruah, Nripsuta Saxena

Agenda

01 Introduction

02 Product Development

03 Demo and Evaluation

04 Conclusions

Human Affect

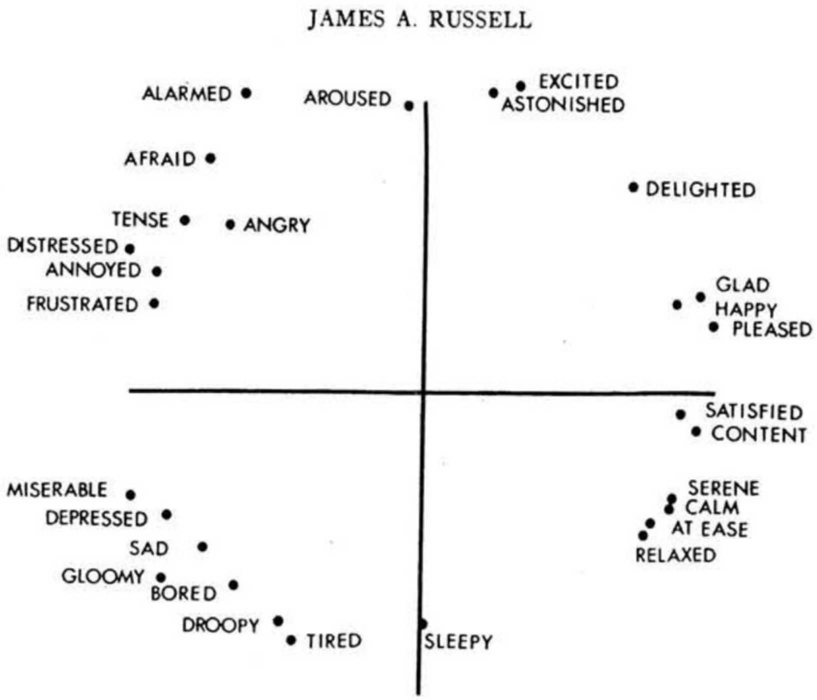


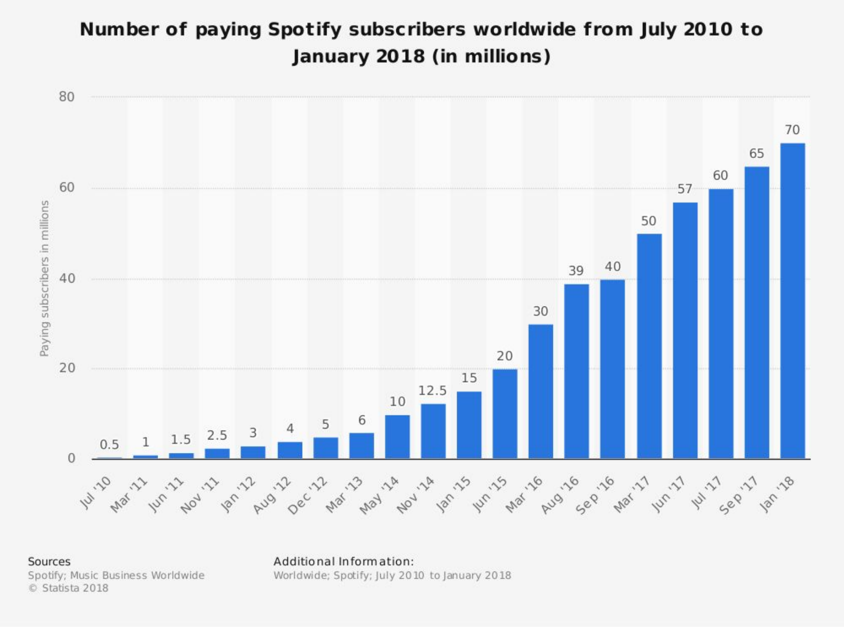
Figure 3. Multidimensional scaling solution for 28 affect words.

Music in the Digital Age

30M+

Song catalogue

Spotify Playlists are curated by music experts around the globe or generated based on listening habits.



Music and Affect

Lyrics

Emotion conveyed through text

Audio Signal

Spectrogram, Rhythm, Tone, harmonic change, loudness, roughness

"Music emotion recognition (MER) detects the inherently emotional expression of people for a music clip."[7]

Music Emotion Recognition Methods

Methods

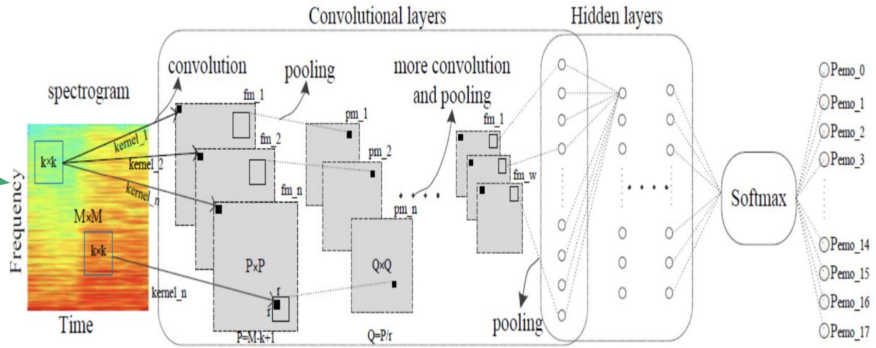
1000 Songs for Emotional Analysis

- SVR
- Bi-Lstm

Liu, et al. (2017)

- Convolutional Neural Network (CNN)

[8]



Project Goal

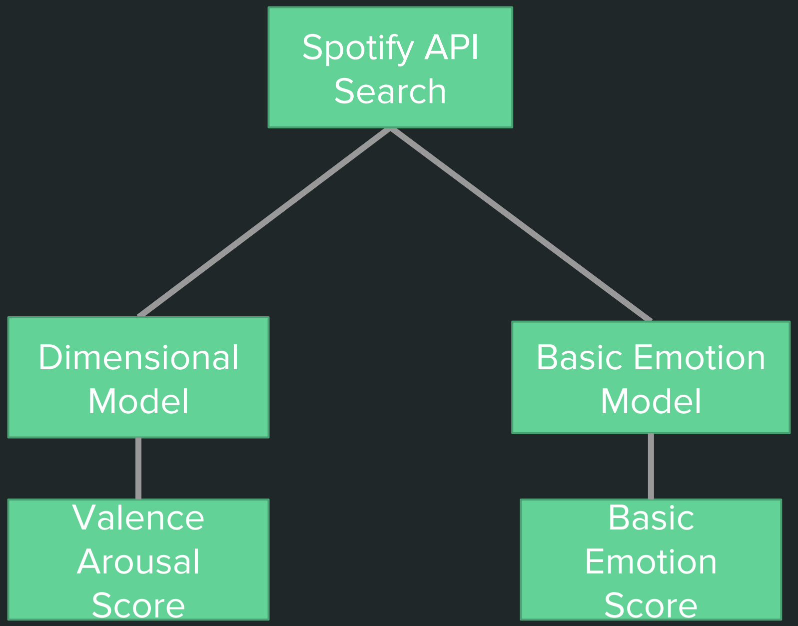
GOAL: Generate mood-based Spotify playlists

1 Use ML techniques to determine the mood of a song

2 Allow users to enter songs that exemplify the mood of the playlist they want and generate a playlist from their library

1.

Use ML techniques to determine the mood of a song

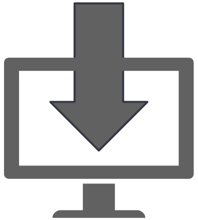


Spotify API

1

- Search for songs
- Get users songs

2



- Download previews for analysis
- Extract song information

3

- Create user playlists

Lyrics API

GENIUS

Annotate the world.

Basic Emotion Model- Dataset

500

Full Songs from Cal500Exp dataset[10]

Segment-level Hard Labels

↓

Song-level labels

Start Time	End Time	Emotion-Angry/Aggressive
164.273	169.655510	1.00000
98.984490	104.491973	1.00000
158.765714	164.273197	1.00000
0.0000000	6.558912	1.00000
169.655510	175.137959	1.00000
54.874558	63.911837	1.00000
39.178231	44.685714	1.00000

Basic Emotions Model-Preprocessing

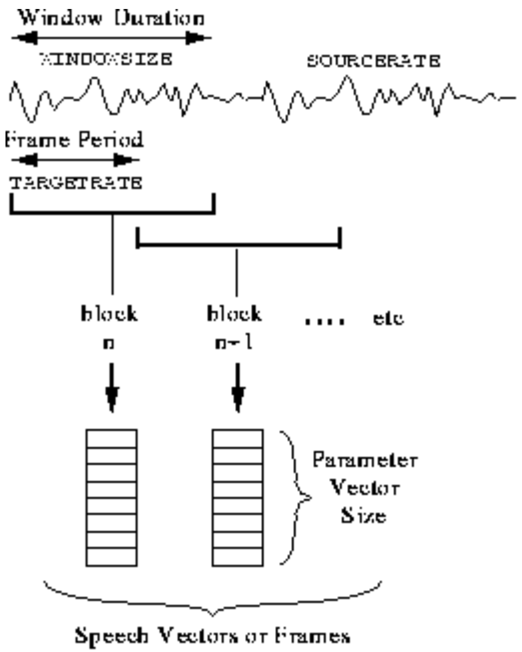


Fig. 5.2 Speech Encoding Process

1 | Resample everything with rate 22.5KHz

2 | Make Spectrogram 257 x 257 by parametrized N_overlap[11]

3 | Retrieve Spectrograms

Simplicity

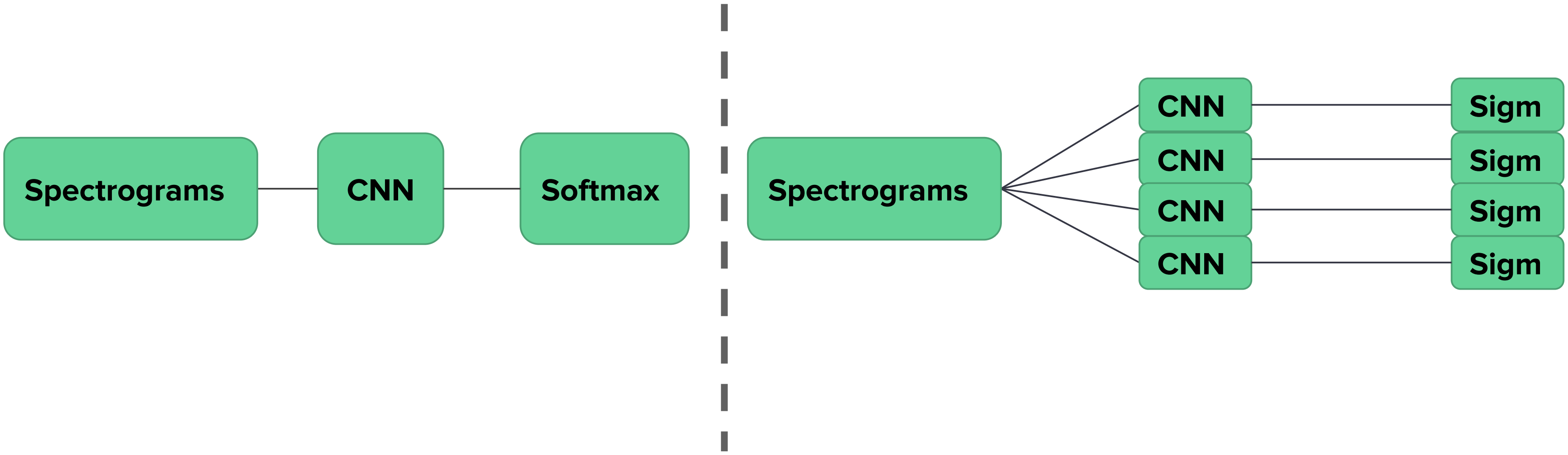
Expect to learn frequency-specific and time-specific patterns

Drawback: repeated segments for shorter songs, lose info for longer songs

Basic Emotion Model-Model Architecture

Challenge: Multi-label Class competition

Skewed dataset



Basic Emotion Model-Results

Test Accuracy

	Anger	Fear	Joy	Sad
Our Model	78%	68%	70%	96%*
Majority Class	88%	56%	68%	96%

Basic Emotion Model- Drawback and Improvements

Issues

Computing resources

Small dataset

Train on long songs but applied on short songs

Imbalanced class

Improvements

Chunking training/test data for consistency

Text Analysis Basic Emotion

Corpus

NRC Word-Emotion Association Lexicon

- A score of 1 means that the *w* conveys the highest amount of emotion *e*.
- A score of 0 means that the *w* conveys the lowest amount of emotion *e*.

Lyrics Analysis

Scan lyrics for words matching in corpus and get their basic emotion scores

- Sum up total scores for the four basic emotions and divide by the number of words found in the corpus

Output

Vector of probabilities for the four basic emotions scores: anger, fear, joy, sadness

Dimensional Emotion - Dataset

- MediaEval 2013, 2014
  - 1744 songs, 45 second clips at 44.1 kHz
  - Valence - Arousal rating between 1-9
  - MTurk Annotation
- Support Vector Regression

	Arousal		Valence	
	RSME	R^2	RSME	R^2
BLSTM (2013)	0.10	0.59	0.11	0.42
SVR (2013,2014)	0.115	0.493	0.107	0.472

Dimensional Emotion - Features

openSMILE:) by audEERING™ → **260** Features

MFCC	Power Spectrum of the audio signal + mel filter bank
Spectrum	Flux, density, roll-off and flatness of fourier transform
Chromagram	Pitch classes in music, melody
Jitter	Variation in periodicity
RMS Energy	Amplitude

Dimensional Emotion - Features

openSMILE:)
by audEERING™

260
Features

MFCC	Power Spectrum of the audio signal + mel filter bank
Spectrum	Flux, density, roll-off and flatness of fourier transform
Chromagram	Pitch classes in music, melody
Jitter	Variation in periodicity
RMS Energy	Amplitude

Text Analysis Dimensional Model

Corpus → **Lyrics Analysis** → **Output**

Corpus

NRC VAD Lexicon

- Valence-Arousal scores of each word in corpus

Lyrics Analysis

- Scan lyrics for words matching in corpus and get their valence arousal score
- Sum up total valence and arousal and divide by the number of words in found in the corpus

Output

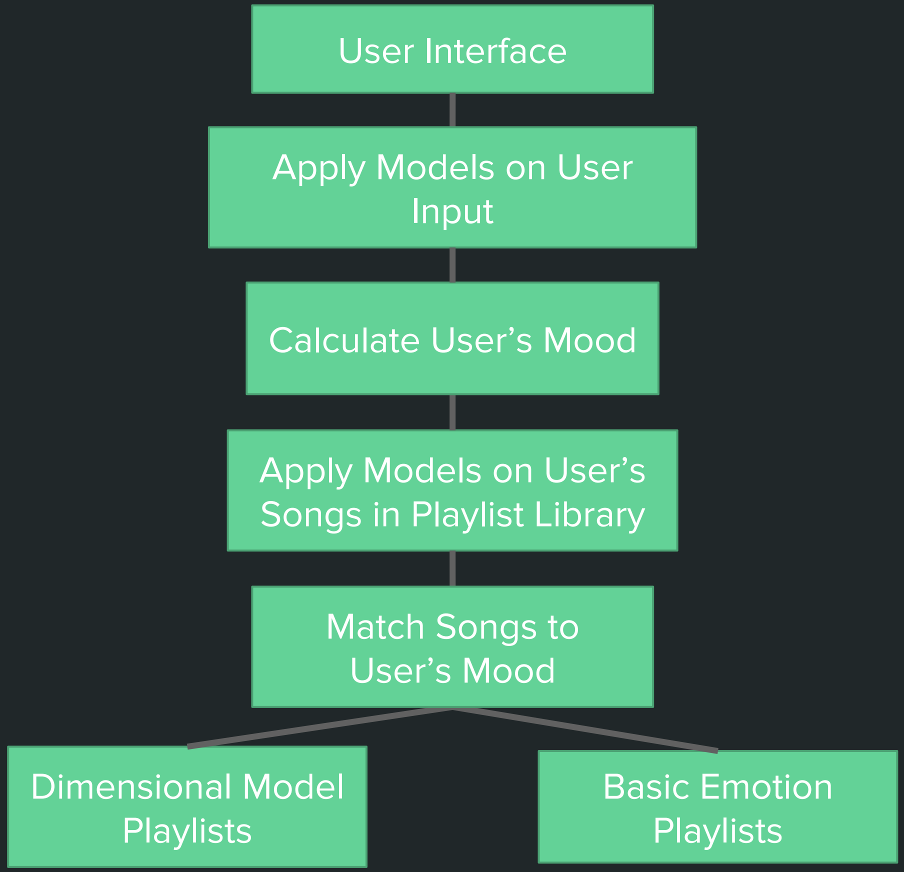
Valence - Arousal scores for each song

Outputs

Basic Emotion	Audio: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$ Text: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$
Dimensional	$[x_{text\text{-}arousal}, x_{text\text{-}valence}, x_{audio\text{-}arousal}, x_{audio\text{-}valence}]$

2.

Allow users to enter songs that exemplify the mood of the playlist they want and generate playlists from their library



User Input

1

Ask user to log on to Spotify

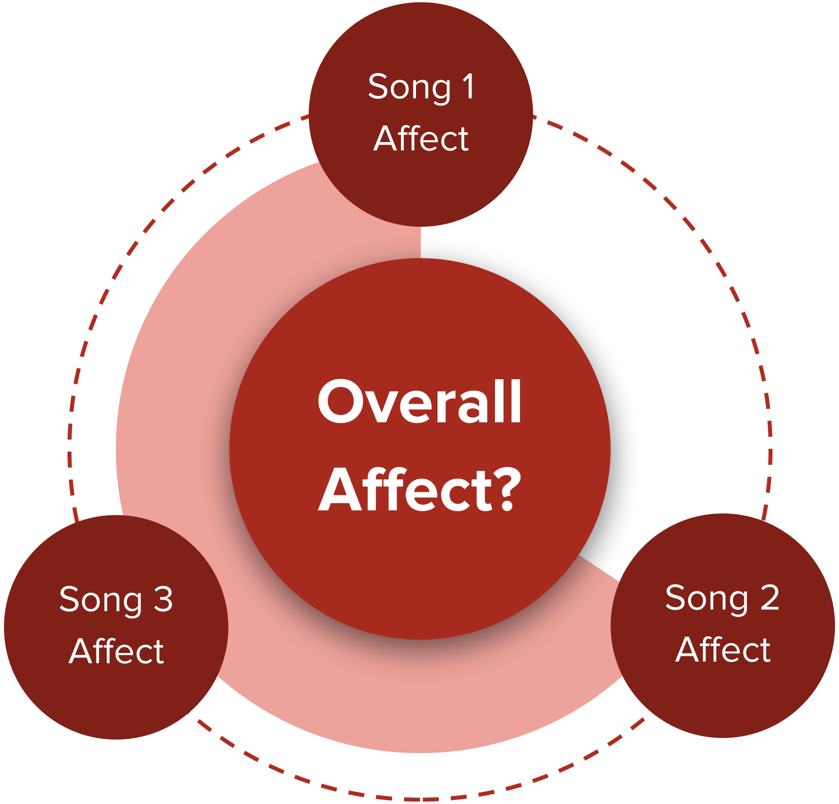
2

User fills out 3 songs that match their current mood in user interface

3

User input is passed to Spotify API to extract song information for the models

Determining Affect Based on User Input



Determining Affect Based on User Input

	Basic Emotion	Dimensional
Averaging (over 3 songs)	Audio: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$ Text : $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$	$[x_{text\text{-}arousal}, x_{text\text{-}valence}, x_{audio\text{-}arousal}, x_{audio\text{-}valence}]$
Clustering	Audio: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$ Text: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$ Audio: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$ Text: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$ Audio: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$ Text: $[p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$	$[x_{text\text{-}arousal}, x_{text\text{-}valence}, x_{audio\text{-}arousal}, x_{audio\text{-}valence}]$ $[x_{text\text{-}arousal}, x_{text\text{-}valence}, x_{audio\text{-}arousal}, x_{audio\text{-}valence}]$ $[x_{text\text{-}arousal}, x_{text\text{-}valence}, x_{audio\text{-}arousal}, x_{audio\text{-}valence}]$

Choosing Songs for Dimensional Playlist

Metric

LET: **Target vector:** $t = [t_{text-arousal}, t_{text-valence}, t_{audio-arousal}, t_{audio-valence}]$

Song vector: $s = [s_{text-arousal}, s_{text-valence}, s_{audio-arousal}, s_{audio-valence}]$

Normalized Euclidean Distance:

$$d(t,s) = \begin{cases} \frac{\sqrt{\sum_i^2 (t_i - s_i)^2}}{\sqrt{2}} & \text{If t or s has only audio scores or text scores only} \\ \frac{\sqrt{\sum_i^4 (t_i - s_i)^2}}{\sqrt{4}} & \text{If t and s has both audio and text scores} \\ 0 & \text{Otherwise} \end{cases}$$

Choosing Songs for Basic Emotion Playlist

Metric

LET: Target vector: $t_{audio} = [p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$

$t_{text} = [p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$

Song vector: $s_{audio} = [p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$

$s_{text} = [p_{anger}, p_{fear}, p_{joy}, p_{sadness}]$

Total Variance Distance of Probability Measures:

$$d(T,S) = \begin{cases} sup_{emotions}|T_{text} - S_{text}| & \text{If only text probabilities exist in both T and S} \\ sup_{emotions}|T_{audio} - S_{audio}| & \text{If only audio probabilities exist in both T and S} \\ \frac{sup_{emotions}|T_{audio} - S_{audio}|}{2} + \frac{sup_{emotions}|T_{text} - S_{text}|}{2} & \text{If both text and audio probabilities exist in both T and S} \\ 0 & \text{otherwise} \end{cases}$$

Averaging

Find k songs with min distance when comparing target vector to song vector

Clustering

For each of the three target vectors: Find k/3 songs with min distance when comparing each target vector to song vector

Output

Playlist 1	• Dimensional Model • Averaging Method
Playlist 2	• Dimensional Model • Clustering Method
Playlist 3	• Basic Emotion Model • Averaging Method
Playlist 4	• Basic Emotion Model • Clustering Method

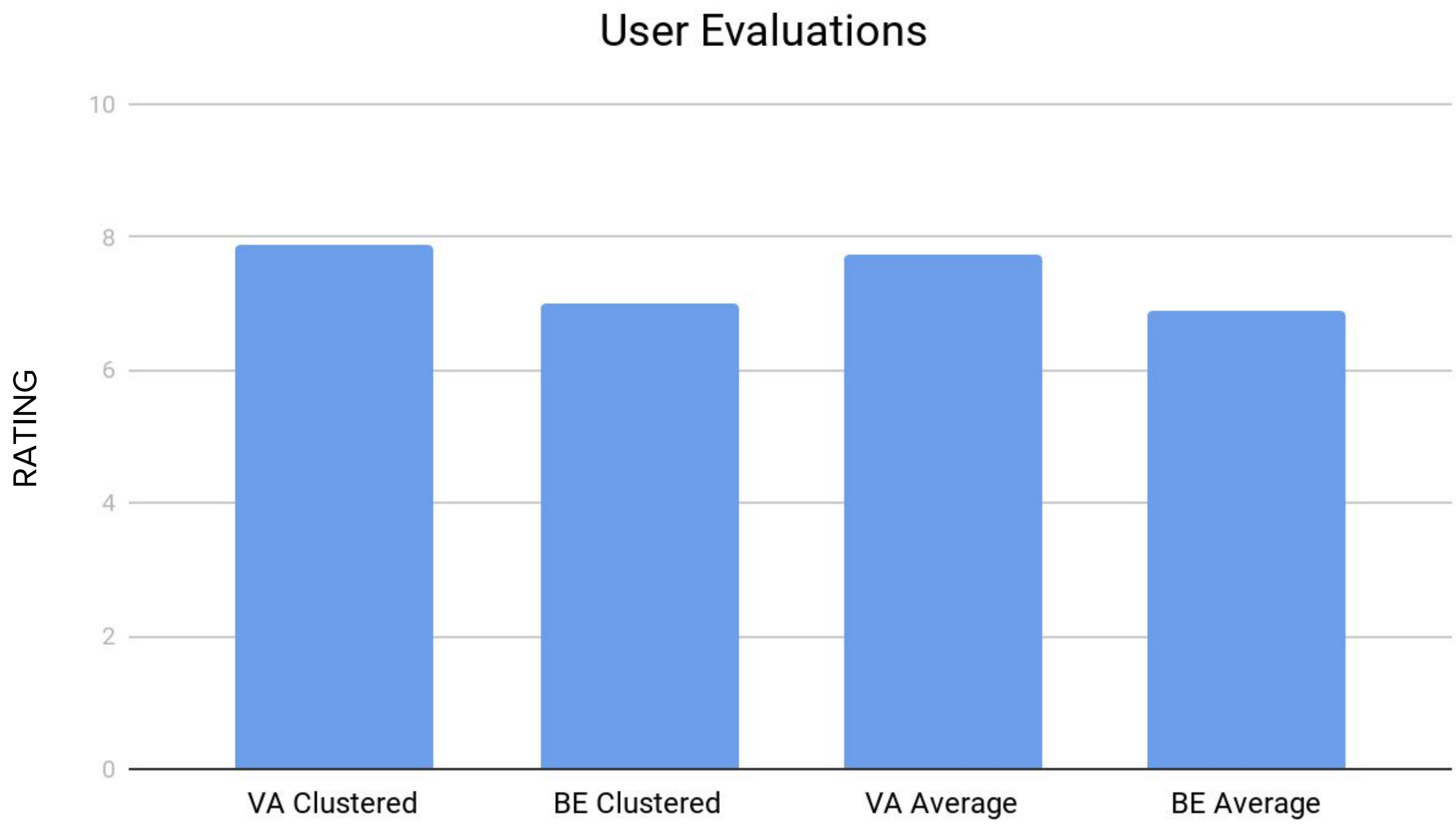
20

Songs Max. Per Playlist

Includes the user input songs

DEMO

User Testing


User Evaluations
RATING
10
8
6
4
2
0
VA Clustered
BE Clustered
VA Average
BE Average

Conclusions

- All models did a decent job building playlists
- Fear and joy were more learnable than anger and sadness
- For Valence-Arousal, mel-frequency cepstral coefficients (MFCC) are still the golden feature
- Training models are sensitive to length

Limitations

- Spotify doesn't have mp3s for all songs
- Must have user's Spotify login before testing the app
- API can't find lyrics for all songs
- Genius API is slow

- Handles English songs only

Future Work and Implications of this Project

1 Improvements

- Design better ways for evaluation
- Incorporate location into playlist generation, potentially improving the playlist quality

2 Future Work

- Add genre functionality
- Conduct more thorough analysis comparing the basic emotions model and the valence arousal model
- Include non-English songs

3 Implications

- Use model to improve Spotify's shuffle
- Use the tool for mood regulation

Division of Labor

01	Christine Chao	• Project Manager • Model for lyric analysis • Front End Development
02	Sophia Mallaro	• Spotify API integration • Playlist generation
03	Sabyasachee Baruah	• Audio feature extraction • VA Model for audio signals
04	Nripsuta Saxena	• Researched questionnaires for determining affect • Front end development
05	Ming Chang Chiu	• Audio feature extraction • Basic Model for audio signals

Thank You! Questions?

Sources

[1] Inskip C, Macfarlane A, Rafferty P. Towards the disintermediation of creative music search: analysing queries to determine important facets. International Journal on Digital Libraries. 2012;12(2):137–147.
[2] Aljanaki A, Yang Y-H, Soleymani M (2017) Developing a benchmark for emotional analysis of music. PLoS ONE 12(3): e0173392. doi:10.1371/journal.pone.0173392
[3] Yang, Yi-Hsuan, and Homer H. Chen. "Machine recognition of music emotion: A review." *ACM Transactions on Intelligent Systems and Technology (TIST)* 3.3 (2012): 40.
[4] Ferguson, Yuna L., and Kennon M. Sheldon. "Trying to be happier really can work: Two experimental studies." *The Journal of Positive Psychology* 8.1 (2013): 23-33.
[5] Xin Liu, Qingcai Chen, Xiangping Wu, Yan Liu, Yang Liu. “CNN BASED MUSIC EMOTION CLASSIFICATION”
[6] Huang, Y.-S., et al. (2018). Pop Music Highlighter: Marking the Emotion Keypoints. Transactions of the International Society for Music Information Retrieval, 1(1), pp. 68–78. DOI: https://doi.org/10.5334/tismir.14
[7]Lin, Chingshun, et al. "Music emotion recognition based on two-level support vector classification." *Machine Learning and Cybernetics (ICMLC), 2016 International Conference on*. Vol. 1. IEEE, 2016.
[8] Huang, Y.-S., Chou, S.-Y., & Yang, Y.-H. (2017). Music thumbnailing via neural attention modeling of music emotion. In: *Proceedings of the Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 347–350.
[9] Harmon-Jones, Cindy, Brock Bastian, and Eddie Harmon-Jones. “The Discrete Emotions Questionnaire: A New Tool for Measuring State Self-Reported Emotions.” Ed. André Aleman. *PLoS ONE* 11.8 (2016): e0159915. *PMC*. Web. 3 Oct. 2018.
[10] S.-Y. Wang, J.-C. Wang, Y.-H. Yang, and H.-M. Wang, “Towards Time-varying Music Auto-tagging Based on CAL500 Expansion,” in *Proceedings of IEEE International Conference on Multimedia and Expo*, Barcelona, Spain, 2014.
[11]X. Liu, Q. Chen, X. Wu, Y. Liu, and Y. Liu, “CNN based music emotion classification” arXiv preprint arXiv:1704.05665v1, 2017.