Music Generation and Song Popularity Prediction using Artificial Intelligence - An Overview

Vaishali Jabade, PhD
Electronics and
Telecommunication,
Vishwakarma Institute of
Technology,
Bibwewadi, Pune, India

Vedang Deshpande
Electronics and
Telecommunication,
Vishwakarma Institute of
Technology,
Bibwewadi, Pune, India

K. Aditya Kumar
Electronics and
Telecommunication,
Vishwakarma Institute of
Technology,
Bibwewadi, Pune, India

ABSTRACT

As the musical industry is rapidly growing, there is an increasing demand for digital platforms for production and consumption of music. With this digitization, a lot of data regarding artists and tracks is available for analysis. Since music production is also digitized, methods for automating this process are emerging as well. The goal of this paper is to explore the methods of generation and popularity prediction. This will benefit , both the creators(music producers, music directors, arrangers, sound engineers) and also the business personnel (Artists and Repertoire, Record labels, artist managers, music distributors and streaming services). Music generation is the process of composing, and arranging melodies(composed of musical notes, within the restrictions of music theory). The popularity of a song depends on various factors such as hotness of the artist, tempo, scale, melody, emotion etc.

Keywords

Melody generation, Music popularity, Music theory, Popularity Prediction

1. INTRODUCTION

Music Information Retrieval(MIR) community has proposed a standard definition as ".....the melody is the single (monophonic) pitch sequence that a listener might reproduce if asked to whistle or hum a piece of polyphonic music, and that a listener would recognize as being the 'essence' of that music when heard in comparison".[1] Therefore, melody is a pattern or sequence of musical notes which is musically sensible to the human ear.

Multimedia content items' popularity has been always considered important, because it plays a crucial role in dealing with various issues of content management such as recommendation, search and retrieval, network content caching, advertisement, etc.[2] Thus, there have been significant research efforts to understand and predict content popularity, especially for images and videos.[3] Since the digital sales of music have been on the rise recently, popularity prediction has become more practical due to the availability of the metadata.

The business oriented personnel in the Music industry are really interested in a song's popularity. It can be used in recommending the users the correct songs and also curating the playlists (collection of songs). Also, the popularity can be used to boost (advertise) the potential tracks which have a high popularity probability.[4]The popularity of a song maybe an important factor in generating playlists.[5].

Some basic questions arise when thinking about popularity prediction of tracks[2]:

1. How can the popularity of a song over time be described from a music chart?

2. What are recognizable characteristics of popularity metrics in long-term real-world chart data?

3. To what extent can the popularity metrics of a song be predicted from the audio signal?

4. What are effective audio features showing good prediction performance?

Following answers were found to the above questions:

1. Popularity cannot be described with one value as it is a temporal sequence. Song's popularity keeps on changing over time so considering one value doesn't make much sense.

2. Tempo, Scale of the song, consumer details like locality, song history, listening duration, etc. can be considered as important popularity metrics.

3. Audio signals don't provide much information conducive to the popularity of the track other than scale, BPM, etc. Hence, metadata and other details need to be considered.

4. Last answer will be analyzed meticulously in the further section of the paper.

Due to the internet, there is a boom in the distribution of music, leading to more competition in the market.[7] And also, since, there is ample data available about the nature of these songs according to the above mentioned factors, it is possible for producers to advantage of this and use it to create music.

There are many emerging applications wherein producers can automate the process of creating music to various levels. There are platforms such as (that use artificial Intelligence as a collaborator in the process of making a song. There are also, some applications, as will further be seen in the paper, that can generate custom melodies that are classified according to the mood the artist wants to create.

After creation of music, the track might need to undergo a process of mixing/mastering. Audio mastering is the process of polishing the recorded musical tracks so that it can be listened in any environment. In this process errors are corrected, frequencies adjusted so that no instrument clash against each other harmonically, and 'loudness' and 'sweetening' added through signal processors and effects including, but not limited to, peak limiters, harmonic distortion, maximizers, multiband equalizers and compressors, and exciters: the 'toolchain'. there are also emerging

applications that seek to complete this process with minimal or no human intervention.

2. STEPS INVOLVED

2.1 Steps followed by a Human to generate a melody

Identifying scale / mood of the tune to be composed.

1) Experiment with the melody in practice on an instrument.

2) Layering the melody with more counter-melodies for making the whole musical piece (Layers generally include chords, pads, etc)

2.2 Parameters for popularity [2][3]

2.2.1 Debut

This is defined as the rank score of a song when the song appears first in a chart. It indicates the initial popularity of the song.

2.2.2 Max

This is defined as the maximum rank score of a song during the whole period, measuring the maximum popularity of the song.

2.2.3 Mean

This is defined as the average rank score of a song over the whole period during which the song appears in a chart.

2.2.4 Standard deviation

This is the standard deviation of the rank scores of a song over the whole period during which the song appears in a chart. It describes how much the popularity of a song has changed over time.

2.2.5 Length

This is defined as the time period (e.g., the number of weeks) during which a song appears on a chart.It measures how long a song has been popular, and thus can identify steadily popular songs.

2.2.6 Sum:

This is defined as the sum of the rank scores over time. Although this is similar to Length, Length measures only the time period, whereas, Sum considers the rank score during which a song stays in a chart. Thus, it basically describes the overall popularity of a song when the whole time period is considered. Sum is different from Mean because the period during which a song appears in a chart is different for different songs.

2.2.7 Skewness

This is the skewness (i.e., the third moment) of the rank scores of a song. This partially describes the dynamic patterns that a song gains and loses popularity. A positive value of Skewness means that the song has become popular fast, reached its highest popularity, and become unpopular slowly; a negative value of Skewness indicates that the song has become popular slowly, then lost its popularity fast.

2.2.8 Kurtosis

Together with Skewness, the kurtosis of a song describes the patterns of growing and declining popularity. The faster the popularity growth of a song is, the larger the value of Kurtosis is.

2.2.9 Chroma

It describes the instantaneous harmony at a particular moment. The chord prediction algorithm is used for obtaining the probability distribution around the chords.

2.2.10 Rhythm

A 2.97 long Hamming window is applied throughout the song and further through Fourier analysis, fluctuation pattern is determined for a moment in the song which helps in determining the time signature or rhythmic component of the song.

2.2.11 Timbre

Timbral components are mostly determined by the frequency components of the song. Hence Mel-frequency cepstral coefficients (MFCCs) are used for determining this.

Thus using both analytical as well as musical features of the song, this paper helps in making a prediction about whether a song will be successful commercially or not.

2.3 Approaches taken/studied

The approaches studied in this paper can be broadly categorized into three categories, generation, analysis and prediction. They are listed below[7]:

1) Melody generation using Music theory

2) Melody generation using Genetic theory

3) Analysing Melody to predict the emotions in the melody

4) Artificial Intelligence based audio mastering

5) Predicting popularity of a song using Machine learning

3. METHODOLOGY

3.1 Melody Generation System based on Theory of Melody Sequence [8][9]

This approach is built on the implication realization model of Music theory. Works in this field done previously, executed the process by learning the probability model of the transition of pitch and then the next coming pitch is predicted following the melody input given by a user.[10].Another approach is to generate based on evolutionary computation . The main disadvantage in this approach is that music is not labelled/ abstracted , Thus, they were not able to output certain melody, which is rarely observed in the training data[11].

But this can be overcome by the application of IRM because one of fundamental assumptions on which the IRM is based on the fact that listeners of music consciously or unconsciously predicts next melody when listening to music. The IRM and properly abstracts melodies in training data. It analyses and classifies the melody in this abstraction process The IRM can analyse atonal and tonal music.

This system which is built on the IRM consists of two models; that of symbol sequence transition and that of generating tones from symbols. With the former model, the symbol transition probability model is trained with the results of the IRM analysis. [12]

The system then generates an optimal symbol sequence according to the probability model. Then, from a set of tones, each symbol sequence generates a melody

Two training songs were taken and it was trained for 15,60 and 100 times. Two types of evaluation were carried out which sought to answer the following questions[13].

(a) An the generated melody be evaluated by IRM

(b) Does the generated melody have elements of the training songs

The findings were subjected to subjective human Judgement, asked 15 subjects to rate the similarity of the generated melody with the training song, on a 5 point basis.

This experimentation confirmed that the generated melodies can be analyzed according to the IRM and also that the generated melody better reflects the tone sequences of the original song as the number of training increases. It verified that elements contained in the generated melody is the same as that of the training song. The proposed system was evaluated through subjective human judgment and the results showed that the system properly generated melodies.

3.2 Melody Generation based on genetic algorithm

One way to compose a melody from a previous given harmonic sequence. A similar technique is used by musicians when they freestyle or improvise off the dome. To state in other words, when there is a predefined sequence of chords and the melody created must fit following theoretical rules of modern music. This particular method exploits the above principle.

In this Algorithm, (eg. mutation, selection, crossover, etc.) child melodies are generated from these new melodies for maintaining genetic diversity[14].

1.Initial Population Generation: Initial individuals are generated randomly based on features like rhythm, pitch, etc. in each mellody block.

2.Fitness Calculation: In the proposed system, the fitness of the gene is calculated based on:

(i) transition of rhythm

(ii) transition of the number of sounds per bar

(iii) similarity of rhythms between phrases

(iv) difference between two sounds

(v) length of rests

(vi) rate of unique sounds of scales

(vii) rate of unique sounds of chord

3.Selection : Based on fitness calculated, individuals are selected based on roulette selection.

4.Crossover : New individuals (melodies) are generated from the selected melodies in step 3 by the multi-point crossover.

5.Mutation : For maintaining genetic diversity, mutation (i.e. alteration in properties of the melody) is carried out.

6. Repeat : Repeat steps 2-5 till a given time Tmax.

In the proposed system, for keeping the generated melodies diverse including notes and rests whose length and position are unnatural, a crossover point is chosen from places other than the middle of a triplet or a rest.

This algorithm provides an apt way of generating "natural/musically correct" melodies as it analyzes the melodies structure wise and also chords wise[15]. Also, the genetic algorithm makes sure that there's diversity maintained in the selection of the generated melodies and also mutation is carried out when needed[16][17].

This algorithm added the implementation more music theory to the previously existing genetic algorithm to improve results. The mutations operators improved the phrases and gave diversity to the composition. This resulted in obtaining great melodies in a good looking interface. This can be used as a reference for beginners who are learning musical improvisation.

3.3 Analysing Melody to predict the emotions in the melody

This algorithm is important as consumers consume music mood-wise, a mechanism to label music databases according to this parameters is needed. This can help composers to judge if the emotion they want to convey is really being reached. this can help in building a melody generator, that composes according to the mood requirement[18]

When it comes to measuring music emotion two things are significant - what emotion is aroused,(discrete) and how hyping or calming, is it(continuous)?

Emotions[19] studied using this method

I. Calm
II. Happy
III. Excited
IV. Sad
V. Heroic

To analyse the melody, the melody must first be extracted from the polyphonic music. There are various methods to do this:

[1] Goto's method
Probability density function of input audio signal is plotted and the melody line using F0's probability density function is observed.

The method is capable of extracting melody from polyphonic sounds irrespective of the source. Disadvantage of this method is must contain a clear lead singing voice or instrument tone. There is also a problem identifying a set of candidate pitches for the melody and identifying which of those identified pitches are related with the melody.

[2] Paiva's Method
This method uses Multi pitch Detection(MPD) process to identify the set of candidate pitches. This involves identifying the pitch values related with the melody by sequentially carrying out a stable pitch, a identifying step which indicate the presence of musical notes and a step for eliminating irrelevant musical note.

[3] Matija Marolt's approach
This approach for extracting melody is based on timbre that is similarity of melodic fragments

The approach consist of the basic steps as;

a) finding predominant pitches using Goto's method,

b) tracking and merging continuous pitch contours for forming melody fragments,

c) finding fragments with well-defined melody using some predefined assumptions,

d) forming the final melody using the highly dominant

clusters of melody contours identified in the previous steps

[4]   Justin Salamon and Emilia Gomez.
This study also is based on the fact that the melody of a music file can be obtained by extracting the fundamental frequency (f0) values. Their approach is based on the creation and characterization of pitch contours (time continuous sequences of pitch candidates) grouped using auditory streaming cues.

By defining a set of contour characteristics, they have defined a set of rules to filter the melodic contours from non-melodic contours of the obtained whole pitch contour.

After Fixing the method for melody extraction as the Justin Salamon and Emilia Gomez's, data set preparation was done . 90 seconds fixed as the standard length. Chorus and first verse covered so all unique sounds covered and repetitive sounds are not.

Each and every song has a melody line, which consist of a series of melody pitch contours. For each contour extracted from the above mentioned method, the starting and the ending pitch values were retained, the highest & the lowest pitch values, the mean pitch value, the standard deviation of the pitch values of the contour, the skewness of the pitch values of the contour, the kurtosis of the pitch values of the contour and the range of the pitch of the contour.

 27 features were calculated as described above. In the initial experiments , the system did not perform well enough to classify songs into different emotions, but then , it was found that melody was an important in predicting the emotion. Further experiments which also include non melodic features like rhythm and tempo, showed that this combination perform much better to provide a proper classification.

3.4 AI based Mastering
Mastering is the final polishing stage that the song undergoes so that it can be heard in any environment. This is considered in this paper because interested in the application of Jingle production using AI based melody generation, is also a topic of interest. For Jingle production, the final product needs to be polished and well-mastered.

Mastering is really an intuitive process, so automating it is a bit of a challenge. But according to [20], "LANDR" is an online platform which has been successful in automating this process.
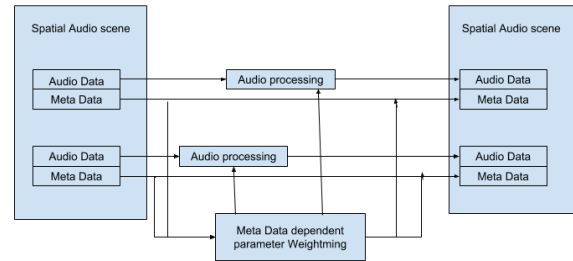

Fig 1: Block Diagram for Mastering

The workflow of LANDR is basically:

1) Digital sum of stems: Stems are basically separate tracks (audio files) for each instrument in the original audio file.

2) Digital Toolchain : This includes equalization (frequency analysis and manipulation), compression, maximizing, distortion, etc.

3) Metadata: Adding metadata to the generated audio file like artist name, genre, etc.

The second step includes feedback from the artist too, as he/she might have some specific changes needed to be made in the master which are taken in to consideration as well.

One conclusion to draw is that AI in the cultural industry of audio mastering will need to strive toward human-centred algorithm design, encompassing both critical listening and creativity, in collaboration with humans rather than through attempts to replace them.

3.5 Methods of Predicting a song's popularity[21]
3.5.1   Ground Truth
 Since no publicly available databases of hits and non hits could be found, data can be scraped data from the Billboard site (Billboard is a magazine which has charts of hit songs for various genres). This site lists all songs which reached the No. 1 ranking in either the United States.

3.5.2   Acoustic Data
This uses an in-house database of approximately 18,500 songs as acoustic data. This data, was obtained by pooling the personal music collections of several members of staff at Hewlett Packard's Cambridge Research Lab. The collection covers many genres ranging from Reggae to Classical although rock songs from the majority of the collection, totaling around 13,000 songs.

3.5.3   Lyric Data
Lyrics are much more easily obtained than audio data as several lyrics repositories are freely available on the internet. However many of these are not standardized or in a format conducive to automated retrieval. Additionally, many of these sites are not comprehensive enough for an effective database.

One site with standardized pages is the Astraweb Lyrics Search site uses data from this site in their experiments. In total they downloaded lyrics for about 500 artists, totaling around 47000 songs, although some songs were repeated on different albums. They als stripped all HTML tags, advertising, and excess information to obtain raw lyrics for the songs.
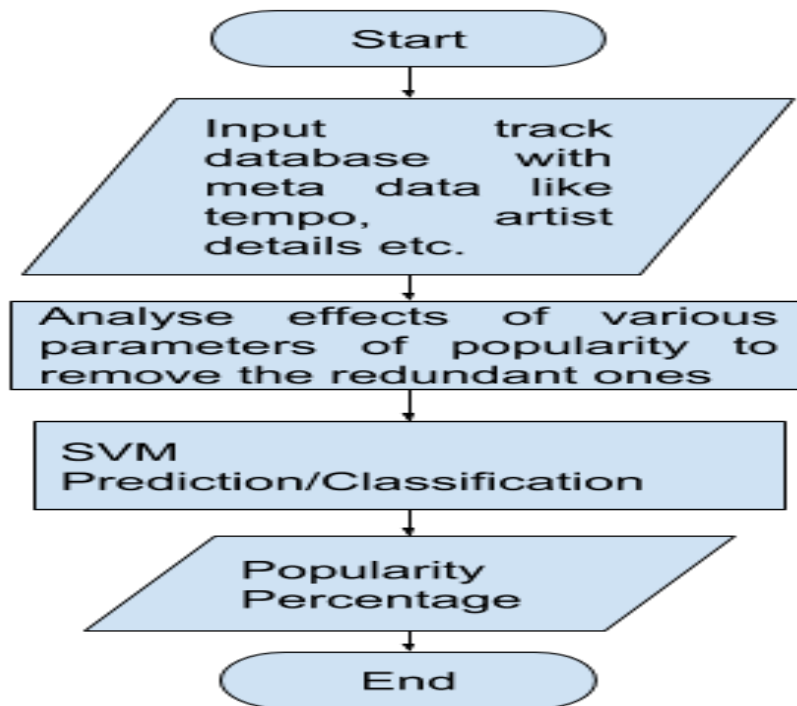
Fig 2: Flowchart for Song Popularity Prediction

4. RESULTS

Table 1 : Correlation Matrix depicting the correlation values of all pairs of parameters in the dataset

	artist familiarity	artist hotness	artist latitude	song hotness	scale	loudness	tempo	time signature
artist familiarity	1	0.812	0.0606	0.4099	-0.033	0.239	0.0635	0.0563
artist hotness	0.812	1	0.041	0.372	-0.02	0.1874	0.0502	0.043
artist latitude	0.0606	0.041	1	0.022	-0.0457	-0.0525	-0.0258	-0.0342
song hotness	0.4099	0.372	0.022	1	-0.0178	0.1865	0.0548	0.044
scale	-0.334	-0.2	0.0457	-0.018	1	-0.0461	-0.0077	-0.0525
loudness	0.239	0.187	-0.0526	0.1865	-0.0460	1	0.1793	0.1161
tempo	0.063	0.05	-0.0258	0.0548	-0.0077	0.179	1	0.0055
time signature	0.0562	0.043	-0.0346	0.0442	-0.0525	0.116	0.0552	1

Table 1 shows the correlation between all pairs of parameters in the song popularity prediction dataset. The parameter with the highest correlation with song hotness is artist familiarity. The data was visualized using bar plots and the following results were obtained.
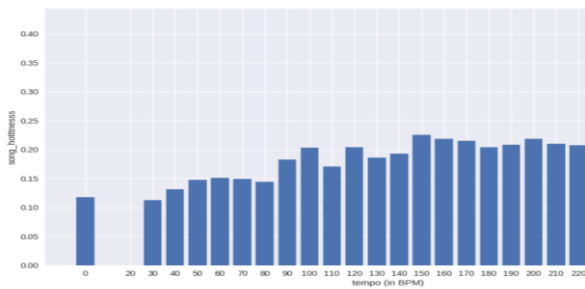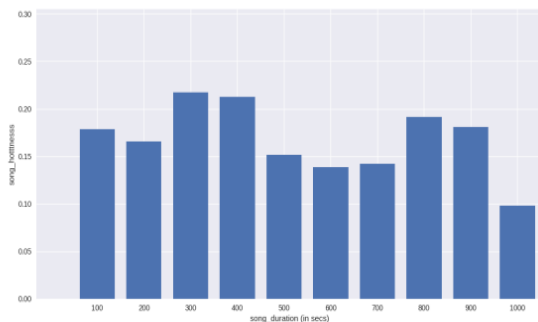

Fig 3: Song Hotness Vs Tempo

Fig. 4: Song Hotness Vs Song duration

From the bar plots(shown in the fig 3,4,5) used to visualize the data, it was found that

(i)   The duration of most popular songs range from 150-300 seconds

(ii) The tempo of most of the popular lie between 100-150 bpm.

The table 3 shows the correlation between all the parameters used to predict the song's popularity.

5. CONCLUSION

Thus, in this paper, the various ways in which the digitally available data can be used to enhance the production, distribution and sales of music in the industry, has been seen. The music theory can be put to use to generate melodies, with respect to reference songs. The methods of generating melodies are elaborated in Table 2. The methods for predicting the popularity are presented in Table 3.

Table 1: Methods for generating melody

Using melody sequence theory	Using Genetic Algorithm
Based on the Implementation-Realization model of music theory.	Based on Genetic model; creating more melodies from a given melody, similar to how artists improvise.

The symbol transition probability model is trained with the results of the IRM analysis. The system then generates an optimal symbol sequence according to the probability model. Then, from a set of tones, each symbol sequence generates a melody	The melody is segregated into verse, chorus, etc. Then each block is analysed on the basis of pitch, rhythm and chord progression. and then suing mathematical models based on processes of genetic evolution, new melodies are created
IRM properly abstracts and labels the data.	Music is not labelled/abstracte

Table 3: methods of predicting song popularity

Ground Truth	Acoustic Data	Lyric Data
Source of data is billboard charts and Meta data of the songs.	Source of data is is the waveforms of songs.	Source are data is just text obtained from the web
Data is readily available, if not it can be scraped and calculated from billboard charts and the scope for computation is high as there are many parameters involved.	Data is most difficult to obtain but the scope for computation is highest	Standardized data is most easily available but the scope for computation is very limited.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Motoki Kikuchi and Yuko Osana, "Automatic Melody Generation considering Chord Progression by Genetic Algorithm", 2014 Sixth World Congress on Nature and Biologically Inspired Computing (NaBIC)

[2] Junghyuk Lee and Jong-Seok Lee, "Music Popularity: Metrics, Characteristics, and Audio-based Prediction", Journal of LATEX Class Files, VOL. 14, NO. 8, August 2015

[3] S. Traverso, M. Ahmed, M. Garetto, P. Giaccone, E. Leonardi, and S. Niccolini, "Unravelling the impact of temporal and geographical locality in content caching systems," IEEE Transactions on Multimedia, vol. 17, no. 10, pp. 1839–1854, Oct. 2015.

[4] B. Logan, "Music recommendation from song sets." in Proceedings of International Society for Music Information Retrieval Conference, 2004, pp. 425–428.

[5] J.-J. Aucouturier and F. Pachet, "Scaling up music playlist generation," in Proceedings of the IEEE International Conference on Multimedia and Expo, vol. 1, 2002, pp. 105–108

[6] J. P. Friedlander, "News and notes on 2013 RIAA music industry shipment and revenue statistics," Recording Industry Association of America, Tech. Rep., 2014

[7] J. Salamon, E. Gomez, D. P. W. Ellis, and G. Richard, "Melody Extraction from Polyphonic Music Signals: Approaches, applications, and challenges," IEEE Signal Process. Mag., vol. 31, no. 2, pp. 118–134, Mar. 2014.

[8] C. D. Manning, H. Schutze: Foundations of Statistical Natural Language Processing, MIT Press, 1999

[9] M.G. Viraj Lakshitha and K.L. Jayaratne, "Melody Analysis for Prediction of the Emotions Conveyed by Sinhala Songs", 2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)

[10] M. Takano and Y. Osana : "Automatic composition system using genetic algorithm and N-gram model considering melody blocks," Proceedings of IEEE Congress on Evolutionary Computation, Brisbane, 2012.

[11] D. E. Goldberg: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Longman Publishing, 1989

[12] B. Han, S. Ho, R. Dannenberg, and E. Hwang, "Smers: Music emotion recognition using support vector regression," 2009

[13] M. Schoen and E. Gatewood, "The mood effects of music," Eff.Music, 1927.

[14] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," Audio, Speech, Lang. …, 2012.

[15] X. Hu, "Music and mood: Where theory and reality meet," 2010.

[16] R. Paiva, "An algorithm for melody detection in polyphonic recordings," Proc. 2nd Music Inf. Retr. …, 2005.

[17] J. Salamon, B. Rocha, and E. Gómez, "Musical genre classification using melody features extracted from polyphonic music signals," Acoust. Speech Signal …, 2012

[18] G. Poliner and D. Ellis, "Melody transcription from music audio: Approaches and evaluation," Audio, Speech, 2007

[19] B. Rocha, R. Panda, and R. Paiva, "Music Emotion Recognition: The Importance of Melodic Features," 5th Int. Work., 2013.

[20] Thomas Birtchnell, "Listening without ears: Artificial intelligence in audio mastering", Big Data & Society. July–December 2018: 1–16

[21] S. D. Roy, T. Mei, W. Zeng, and S. Li, "Towards cross-domain learning for social video popularity prediction," IEEE Transactions on Multimedia, vol. 15, no. 6, pp. 1255–1267, Oct. 2013.