Novel Approach for Music search using music contents and human perception

Velankar Makarand
Faculty Information Technology Department
Cummins College of Engineering
Pune, India.
makarand_v@rediffmail.com

Dr. Sahasrabuddhe H.V.
Formerly Professor KReSIT
Indian Institute of Technology
Mumbai, India
hvs_buddhe@hotmail.com

***Abstract*— Music similarity can be perceived by the listeners in different ways depending on individual preferences to musical parameters. Metadata based similarity measures are used by many websites for music recommendation and retrieval. Metadata attached to music files as text generally include title, albums, genre, artist etc. The main limitation of this method is that it can only work if metadata is available with the music. It may produce wrong results if incorrect metadata is associated with the music file. Present Meta data based systems also lack in analyzing contents. We have conducted experiments with Indian film songs in pairs and triplets to understand listener's perception of similarity. We analyzed content based parameters such as voice, mood, instruments used, lyrics, rhythms, tune etc and their relative importance in similarity perception to majority of the listeners. We propose to use both metadata and content based parameters. We worked on few parameters and worked with some algorithms in our prototype. Our findings can be useful for music Information Retrieval System and the concept is applicable to all musical forms. Perception based framework for music similarity is an attempt to model present and future needs of music listeners with option of adjustment for various musical parameters from listener perceptive.**

***Index Terms*—music similarity, music speed, pitch tracking, musical retrieval, music recommendation**

I. INTRODUCTION

Similarity measures for music are useful in various fields such as search, recommendations and clustering. Music similarity measures have 2 main categories as

1. Metadata-based Similarity measures: These measures are used by many websites such as YouTube etc. Metadata attached to music files as text may include title, albums, genre, artist etc. The main limitation of this method is that it can only work if metadata is available with the music. It may produce wrong results if incorrect metadata is associated with the music file.
2. Contents based similarity measures: These measures can be based on various music parameters such as melody, rhythm, tempo etc. This approach extracts the musical features from the musical file using signal processing techniques. Considering the complexity of music signals, content feature extraction may miss useful information or may provide some erroneous information.

We have attempted to address following issues through this paper.

a. How to overcome the limitations of present Metadata based systems?
b. How music similarity and recommendation can be implemented for an individual listener with changing needs, specific interests and diverse perceptions?

We conducted music perception experiments for the selection of content-based similarity features. A similarity framework is developed based on music similarity features. These different features are used by listeners consciously or unconsciously depending on their own perception of music. Although we have worked on film songs the concept can be extended to other musical forms in the future to adopt subjective similarity framework.

Music perception is the concept & understanding of music from listener's perceptive. Music perception differs from person to person and even the same listener may perceive the same musical clip differently at different times. Music perception also differs with respect to the form or genre of music such as film music or classical music etc. Listening experience can alter an individual listener's perception over time. Seasoned listeners and novice listeners are observed to perceive music differently. A single, fixed framework valid to all types of music and all types of listeners may not be possible. The framework should be adaptable to the personalized experience according to individual expectations and future needs.

II. RELATED WORK

Jin Ha Lee [2] has mentioned about user perceptions and different user behaviors on similarity of music. Daniel Wolf, Tillman Weyde [1] focused on need of evaluation from a musicological perspective. Dash and Liu [4] presented a comprehensive survey of general techniques for feature selection in classification tasks. The paper [3] provides a review of interactive element of user interfaces (UI) in music recommendation systems. Perception and cognition models are explained in review [5]. Papers [6], [8], [11] have thrown light on music cognition and interpretation of music. Achyut Godbole [7] has explored song similarity based on similar ragas or note patterns. Raga Mala[9] has explained Indian Raga music in detail. Yi Liu [10] explored on mood extraction from song databases. Music Researchers are working on better user interface, use of content and context information of song and many musicological aspects

978-1-4799-2102-7/14 $31.00 © 2014 IEEE
DOI 10.1109/ICESC.2014.9

to design efficient user friendly music information recommendation and retrieval system.

III. SIMILARITY PERCEPTION EXPERIMENTS

We have initially focused on popular Indian Hindi film music to understand the process of perception and similarity. Hindi film music was chosen because great variety in singers, musical composers, moods of songs, instrumentations, rhythm etc. We used 50 Hindi popular film songs from the era 1950's to 1980's as music clip samples with 100 possible pairings.

These 100 pairs were selected considering similarity of different parameters and possible combinations of them. The parameters selected by us were singer (or singers in case of duets), mood perception from the song which can be happy or sad or calm feel or romantic etc., tempo of the song, Prominent instruments used such as flute or guitar etc. in the song, tune or musical notes pattern such as songs based on similar raga, Lyrics or the words used prominently to express the emotions or convey the meaning, rhythm or beat patterns used by music composers.

Each experiment consisted of 2 songs played in different sequence for different users and users commented on the possible similarity perception between 2 songs. We have taken feedback from about 80 listeners of different age groups and musical background to understand the similarity perception. We have conducted about 200 experiments with different song pairs and noted the similarity perception along with musical background of the listeners. We also conducted the experiments with about 10 triplets with clips named as A, B, C played in different sequence. We asked the listeners to select 2 most similar clips from them and mention the reasons for their selection. We collected 100 Responses from the triplet experiments.

Following table (Table 1) describes the summary of the feedbacks about similarity perception from these 300 sample feedbacks. Listeners were asked to rate the similarity on the scale, select multiple choices for the parameters considered for the similarity or dissimilarity.

Table 1 Listener Feedback

Similarity Parameter	**Responses**
Singer/ Voice	254
Mood/ Expression	188
Tempo of the song	152
Tune/ musical notes pattern	114
Instrumentation used	83
Lyrics/words-	76
Rhythm/beats pattern	42

From these experiments we identified the common important musical parameters considered by many listeners and other unique parameters used by some listeners for similarity perception. We can use general weightage information of different parameters for typical similarity comparison. We observed that the relative importance of common musical parameters was different from person to person and that some parameters were considered important by different groups/sections of listeners. This leads us to the thought on need of personalization in the proposed framework depending on user's choice of parameters.

IV. PROPOSED SIMILARITY FRAMEWORK

The musical parameters information can be represented in different forms such as artist names for singers or albums or song title in text form, tempo in values on scale, tune or rhythmic information in the musical notes patterns etc. We observed that mood is generally perceived from the tempo of the song and in some cases from the lyrics meaning etc. Considering this multidimensional musical information representation in different forms, we need to adopt different methods for similarity mapping of different parameters.

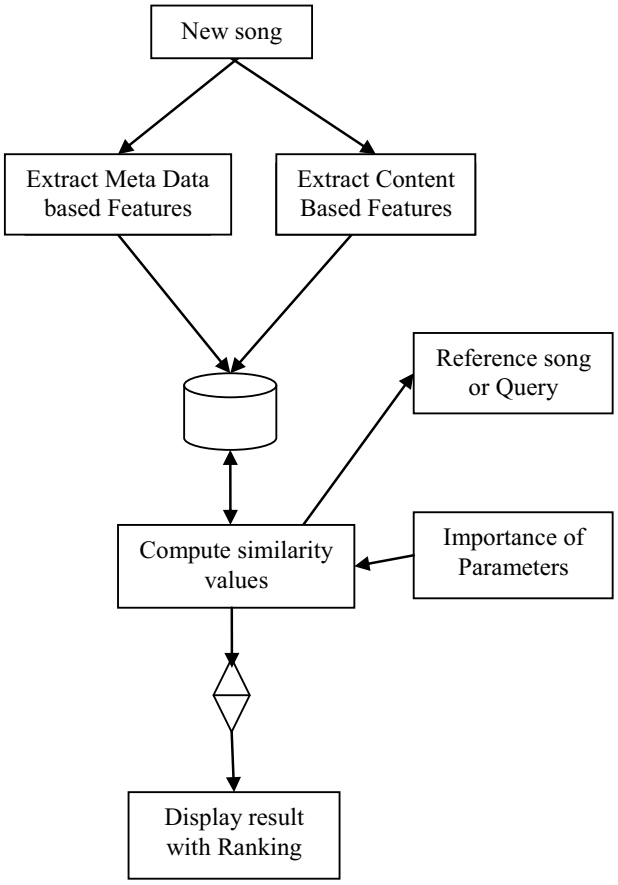

Figure 1 proposed system flow.

For a new song the Meta data based features or keywords and content based parameters are extracted and

stored in the database. User can submit query in traditional form of keywords or select a song in the database or upload a song as a query. For keyword based queries, the proposed system works in similar manner as per the present available systems. For selected song or uploaded song, user can adjust weights according to user's importance to different musical parameters for similarity. After submitting the query, a similarity function computes similarity values on the basis of weightage to different parameters selected by user with reference song. Depending on similarity values computed, the result is sorted according to ranking of possible similar songs.

We propose the similarity function (equation-1) for computing weights for content based parameters. The importance to various musical parameters is flexible for personalization. The initial weights to the musical parameters are fixed based on the music similarity perception experiments. We calculate the similarity value of musical piece with reference to the original music piece using similarity function. We can combine results using metadata or keywords and content based similarity results to model user perception. We believe perception-based approach will give better results compared to a metadata-based approach for search and recommendations. We further believe that our approach will prove superior to a content-based approach which fails to model the typical listener. Our framework will be adaptive to the change in information needs of users.

Similarity function proposed uses various musical parameters with possible adjustable values according to user's parameter adjustments. Standard values will be set according to our perception experiments and general perception from majority of listeners.

$$S [i, j] = \sum^{n}_{m=0} (Fm . Km [i, j]) / n \qquad (1)$$

Where,

Fm- Adjustable factor by user for specific facet m.

Km [i, j] - Similarity matrix value between song i & song j for specific parameter m.

S [i, j] – Similarity matrix value between songs i and j.

n- Number of parameters

Note- Fm, Km [i, j] & S [i, j] all values range from 0 to 1.

The proposed framework can be used in the search system at different stages. User Interface accepts queries as sample musical piece or keywords from the user to find the best possible match along with query personalization option on the basis of selected musical parameters. Query Evaluation Engine will analyze the query with set parameters related to musical parameters. It will calculate the similarity value between different musical pieces with respect to sample piece or keywords and retrieve the musical clips from the databases.

In the proposed search system, we can use the common facet information of each musical clip for clustering and indexing. Ranking can be calculated on the basis of similarity value, listener's web history and other parameters used in text search applicable to music retrieval. We propose the music search system similar to Google search for text retrieval with advanced features as a personalized musical retrieval and recommendation system.

V. SIMILARITY MEASURES MAPPING IN FRAMEWORK

During the similarity experiments, we observed and studied the response of listeners to different musical parameters and how each facet affected their perception of similarity. As an example, tracks featuring singers of the same gender were perceived more similar to each other than tracks featuring one male and one female singer. Different musical parameters need to be represented in a form which can be later used to find similarity in numerical value for each parameter and those individual values can be used for ranking of clips. We need to assign numerical values to each parameter so that they can be used in ranking clips. We found this to be the most challenging aspect of the proposed framework.

For our prototype, we have worked on singer/voice using metadata, mood using speed of music, and note pattern in our present system considering the importance by majority of the listeners during the experiments.

A. Singer/Voice

Singers and album information can be represented with names subject to availability of the information in the metadata. Another possible way can be to store the information as a voice profile. The challenge in this approach is to find and represent voice profile itself for any individual singer and later on matching of the same. Another challenge here is the possible change in individual singer's voice profile depending on artist on the screen or style in different era or mood of song. Identifying and storing the timbre information for possible similarity computing is one of the serious research challenges in music Information Retrieval.

Another possible approach is to store the information as male, female singer or duet etc for the song using bitmaps. Bitmaps is suitable method when a particular parameter can have values from the specific set. Bitmap requires less space for storage and we can use bitwise operations such as EX-OR to find the possible matching.

For example the male, female or duet information of the each song can be represented using 3 bits for each song and one of them can be 1 depending on type of singer. e.g. Bit pattern as 1st bit- male, 2nd bit- female and 3rd bit for duet. This pattern will represent male singer as 100, female singer as 010 etc.

We have used metadata information for singer/album information in our present prototype. Keyword matching is

used for singer/voice parameter for computing similarity value.

B. Mood or expression

Mood or expression of the songs can be possibly represented as different possible feelings or their combinations. Feelings from the song can be happy, sad, romantic, relinquish, devotional, tranquil, disheartened, enjoyable, patriotic etc. Lyrics may play a role here establishing specific mood e.g. words of a lullaby gives the feel of calm and peaceful sleep. Automatic identification of mood is a challenging task and representing same on scale for similarity is even more challenging.

We have focused on 1 mood pair happy/sad in the present prototype. We used speed to represent happy/sad mood. Our decision to use speed parameter for respective mood is based on the experimental results of emotion related experiments carried by us with 13 different emotions for 6 musical clips with sample size of about 100 listeners.

In speech, generally speed is considered as number of words or syllables in unit time. Similarly, music speed is a measure of total number of notes played in unit time. This can also be termed as rate of change of note in unit time. Speed of notes sung plays the role in the tempo perception. Although speed and tempo interprets similar information, they are not same. For some musical clips there is no rhythm accompaniment and we can use speed parameter in such case. For majority of musical clips where rhythm is present, tempo and speed parameters can compliment each other well to reduce possible error in calculations of them.

C. Tempo of the song

We observed tempo of song was major factor for association of mood of a song as exciting/cool. Tempo is the listener's perception of beats per second which can be found out from the actual prominent beats in the song. The more the perceived beats per unit time, the song is perceived as fast and happy song and less the beats, it is perceived as slow and sad song by majority of listeners according to our observations during experiments. Tempo can be majored and represented on the scale and used for similarity association using similarity function.

D. Tune/musical notes pattern

Tune is generally referred to the mukhada or prominent representation of song which is repeated. Automatic tune finding is difficult in the presence of rhythm and other instruments played simultaneously. Tune representation and matching tunes for possible similarity is another difficult challenge.

Notes patterns or melodic phrases are represented using bhatkhande notation does provide information about notes and duration of notes but information about loudness of the notes is absent. Each notes pattern should be ideally represented with frequency, duration and loudness to compare for possible similarity. Change in any parameters affects the possible similarity perception from listener's point of view.

E. Instrumentation used

Prominent instruments used in the song like guitar, flute etc. influences listener's perception for similarity association. We also observed here era of the song does make a significant impact on the instrumentations used. This observation leads us to consider the composition year of the song as another possible important parameter useful for similarity perception.

F. Lyrics

We observed that lyrics or words used prominently in the mukhada of the song does play role in the similarity perception from the listener's point of view. Lyrics information if available can be used for such similarity matching.

G. Rhythm or bit pattern

Many listeners associated typical rhythms used by music composers for similarity perception. We also observed that listeners with musical background such as playing instruments such as tabla tend to notice and associate rhythms used in the song for the similarity perception more than other listeners. Rhythm patterns can be represented using notations and durations between two rhythm beats. Automatic identification and representation of the rhythmic patterns suitable to find similarity is the challenging task.

VI. PARAMETER USED FOR ESTIMATION

We observed few music samples and also discussed with some musicians and researchers working in this field to finalize the decisions about samples per second, window size, musical note frequency, amplitude information and reference note to be used for estimation.

We decided to use samples after every 10 ms for estimation. This enabled us to observe 4 samples in every 40 ms window size or 100 samples per second. This was decided considering minimum time of note played to be noticed by listeners as per perception studies. We have observed the pitch contour during rapid note variations (taan or glide) to notice notes. We decided to average successive samples considering 40 ms time frame. We shifted window size of 40 ms by 1 sample every time i.e. 10 ms to estimate next window information. This was done because we do not know the beginning of small duration note and the information about 40 ms note duration should be recognized by algorithm.

For pitch estimation, we found that successive notes are about 6% apart from each other in the tempered scale and we decided to consider ± 3 % range about note frequency as the frequency of note for our initial estimation. We can change this parameter for fine tuning as we progress.

Another decision was about loudness or amplitude information. We observed and found that for any clip;

maximum amplitude information can be used to find pauses or silence. In majority of musical clips maximum amplitude was about 85 Db and clip portion with less than 55 Db were pauses. We have estimated about 30 Db from the maximum amplitude can be considered as a perceived audio range and all values below that can be considered as pauses or silence. We have divided the amplitude audible range of 30 DB in 6 slots of 5DB each. We represented them with 'a' to 'f' as maximum to minimum for simplicity in comparison. We assign amplitude 's' to denote silence.

In Indian Music, the reference note (Sa) is not a fixed frequency in Hz. It does very according to different performances by different artists and different tuning of instruments used. Generally reference frequency of women singers is higher than men. For instrumental music, it depends on tuning of the instruments and the performer. We are working on methods to find the reference frequency of the performance. We have used fixed frequency scale at present to represent notes related information.

VII. ALGORITHMS USED IN PROTOTYPE

Following algorithms were used to find different musical parameter values for similarity.

A. Algorithm for Speed of music

We considered minimum time span for any note as 40 ms (window size with 4 samples) to be noticed by any seasoned listener. We have shifted window every time by one sample (10ms).

Initialization
  For each subsequent window
    Calculate amp
      For samples 1 to 4 with amp<> 's'
        Pi = i^{th} sample frequency
        If (P1=P2=P3=P4)
        CurrentP= P1
      If currentP = PrevP
        PrevP= currentP
      Else
        PrevP= currentP
    Write note and amplitude information
    Increment Notes count
    Go to next window
End while loop

Speed of music= note count/time duration

B. Generation of note sequence

We have presented the information on time scale with 1 second as a unit separated by bar to present notations as maximum 25 notes per second and display time at the end of each second. We used the same algorithm used for music speed finding and written the output of notes if found with successive 40 ms time frame along with amplitude.

Lower octave notes were presented with underline and upper octave notes with symbol ' followed by note considering spread in maximum 3 octaves. Middle octave notes presented normally without any specific symbol. 110 Hz to 220 Hz, 220 Hz to 440 Hz and 440 Hz to 880 Hz are three respective octaves used.

Sample output pass 1

Sa d| 3 |Ni d Ni c Ni c Ni c Ni c ni d ni d Dh c

In this sample output Sa, Ni, ni Dh are the musical notes, 3 represent end of 3rd second and d or c represents amplitudes.

Sample output pass 2

Sa d| 3 |Ni d Ni c 4 ni d 2 Dh c

We processed pass 1 output to generate better note sequence by removing repeated notes in sequence. Above is the sample output of pass 2 for the same note sequence with 4 and 2 representing repetition of previous note 4 and 2 times respectively.

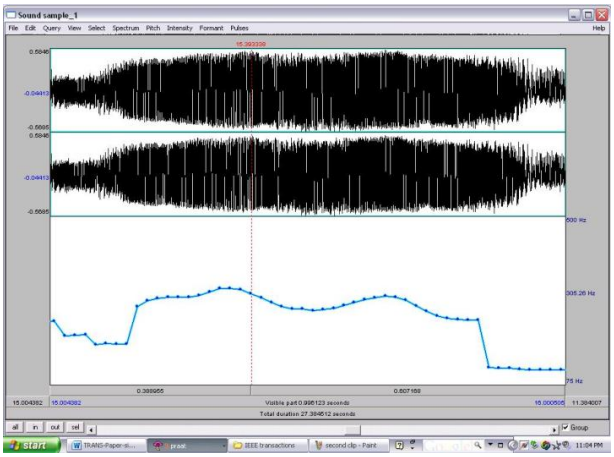

Fig.1- Shows sample waveform and pitch contour for the duration of about one second from a song clip portion.

In the present prototype we used praat software to extract and analyze samples. We worked on these music sample values to get different parameter values for similarity. We came across undefined frequency for few sample instances which was due to possible attack of bits or simultaneous presence of multiple frequencies etc. Preprocessing of audio sample with better pitch tracking algorithm can give better results for pitch estimation and note sequence generation. We are working on applying band pass filters and see the results and also trying to use some other tools for better pitch tracking mechanism.

VIII. GENERATION OF SIMILARITY MATRIX

A. Metadata

For Meta data and keyword information, we used Jaccard's coefficient to calculate similarity value. It represents similarity value in range 0 to 1. We consider common keywords and total number of keywords in any 2 song descriptors in the Meta data.

B. Speed of music

Speed of music is a numerical value for the song or piece of music to represent number of notes changed per unit time. We have computed the value for per second. We computed the difference (d) between 2 different speeds represented for individual songs. The lower the value of d greater is the similarity. Thus for similarity function, we used 1/d value to represent similarity, where we replaced value of d as 1 when difference is 0 to denote max similarity.

C. Tune

Tune similarity is the most challenging task as the tune comparison need to consider with 3 important parameters note (frequency), duration (time) and amplitude (loudness). We are working on possible effective tune matching algorithm. Tune of any song can be represented with most commonly repeating note sequence within the song i.e. generally termed as mukhada of the song. This approach will miss out details about entire song but will reduce computation. In present prototype we focused on note sequence along with its duration and amplitude.

IX. PROTOTYPE RESULTS

We attempted to mainly analyze 3 aspects as Meta data, speed and tune for the first prototype of personalized music search. Subsequent aspects need to be analyzed in more details for completeness of the work. Each musical aspect has its own peculiarity and it is difficult to put it on similarity matrix. We also need to consider music perception studies and music cognition at later stage.

The results for Meta data based similarity for artist/singer information are acceptable but we need to work on many other aspects as possible use of other similarity coefficients, case insensitive comparison, classification of information into possible fields such as album, singers for better comparison etc. This will give us similar results like results from popular music retrievals like YouTube.

The results with speed detection and tune are not as per the expectation and we need to do improvisation in both on many aspects. Speed detection needs to detect proper notes and small pauses correctly as per listener's perception. We are working on perception and better pitch estimation to improve algorithm. Tune similarity needs to take care of songs sung in different scales or different reference notes. Proper identification of reference note is one possible approach or Shifting of note sequence for comparison of tune can be another possible approach. It is a long way ahead to truly realize the proposed search system.

X. CONCLUSION

The proposed music similarity framework based on music perception by listeners can be used on large scale for different types of music in the future. This concept can be effective for music search by users with possible recommendations and can be extended for mobile tune recommendations for mobile users. The proposed framework can adopt the changing needs of users in the future by user preferred music parameters to give them personalized similarity measures.

ACKNOWLEDGMENT

We would like to thank the entire participants for participating the surveys conducted. We are very much grateful to all musicians and computer music savvy people for sparing their valuable time and providing valuable inputs for the success of the work.

REFERENCES

[1] Daniel Wolf, Tillman Weyde- "Adapting Metrics for Music Similarity Using Comparative Ratings" http://ismir2011.ismir.net/papers/PS1-6.pdf

[2] Jin Ha Lee- "How Much Metadata Do We Need in Music Recommendation? A Subjective Evaluation Using Preference Sets" ISIMR Proceedings

[3] P. Åman and L. A. Liikkanen: "A survey of music recommendation aids," Proceedings of the Workshop on Music Recommendation and Discovery, 2010.

[4] M. Dash and H. Liu. "Feature selection for classification". Intelligent Data Analysis, 1(1-4):131–156, 1997.

[5] Pandit Vishnunarayan Bhatkhande (Bhatkhande 1957) Kramik pustak malika-part 1 to 6 Hathras: Sangeet Karyalaya 1st edition, 1957.

[6] Dr. Martin Clayton (Clayton 2001)-"Towards a theory of musical meaning" British Journal of ethnomusicology vol-10/I, 2001

[7] Achyut Godbole, Sulbha Pishvikar (Godbole 2004) "Nadvedh" Rajhauns prakashan, 2004.

[8] Kai Tuuri, Manne-Sakari Mustonen, Antti Pirhonen (Kai Tuuri 2007)-"Same sound – Different meanings: A Novel Scheme for Modes of Listening" Audio Mostly September 27-28 Germany 2007

[9] The Raag-mala music society of Toronto The Language of Indian Art Music Toronto, 2004 (Raag-mala 2004)

[10] Yi Liu, Yue Gao (Yi Liu 2009)- "Acquiring mood information from songs in large music databases"

[11] "Computational models of music perception and cognition :The perceptual and cognitive processing chain" Science Direct Physics of life- (2008) 151–168

[12] Riccardo Miotto and Gert Lanckriet "A Generative Context Model for Semantic Music Annotation and Retrieval" IEEE Transactions on audio, speech and language processing, vol-20, no-4,May 2012.

[13] Myung Jong Kim and Hoirin Kim "Audio-Based Objectionable Content Detection Using Discriminative Transforms of Time-Frequency Dynamics" IEEE Transaction on multimedia vol-14,no 5 Oct 2012.