A Review of Audio Gene Recognition Copyright Protecting Technology[*]

Tan Xinjie[1, 2], Cui Jizhe[1†]

1. College of Economic & Management YANBIAN University, Yanji 133002, China

[†]Email: cuijizhe@foxmail.com

Abstract

The copyright stands the core competence of the cultural industry. The cultural influence depends on the standard copyright management system and supporting information processing technology in the future. This paper set the audio content copyright protection technology as starting point, and introduced the audio gene technology and research background based on characteristics. What's more, an audio gene technology evaluation agency named MIREX established to protect copyright was introduced, which resumed audio emotion clarify standard and related technology. This paper aims to introduce audio content management system based on protect copyright technology and to provide feasible copyright thought for the musicians and enterprises, according to introduce technology research method of audio technology field based on emotion analysis and content analysis.

Keywords: Audio Genetic; Copyright Protection; Technology Research

1 INTRODUCTION

With the popularization of Internet media, audio copyright technology experiences a severe test and become a hotspot of social attention. In 2017, a series of copyright dispute between Tencent Music and Kuwo Music (from Sina news) also illustrates the necessity and urgency of the copyright protection. The infringement of most audio content in these cases is determined by identifying the audio genes of the songs. Audio gene technology discussed in this paper is referred to a Music Information Retrieval Technology aimed to protect copyright based on characteristic audio gene which is extracted from a large number of Music database [1], including digital fingerprint technology characterized by the content as the direction of Music fingerprint identification. Audio gene technology can be widely used in wearable device control, Monitoring and control system and security system, smart home system, logistics field calibration, sound, information identification and interception, music copyright protection, Broadcast Monitoring, etc. And it can be applied to all kinds of intelligent equipment enterprises to identify the music played in the environment, which can be widely applied to automotive music recognition, bar lights switch intelligently according to songs. Based on this, this paper attempts to set audio gene technology as a starting point, summarize the domestic current situation of the development of gene technology from evaluation agency, evaluation principle and evaluation technology, and introduce audio gene technology agency MIREX. Further analysis was based on direction of emotion and content technology research in the field of audio technology, revealed an audio content management system based on copyright protect technology, to provide feasible copyright protection for musicians or enterprises.

2 AUDIO GENETIC PRINCIPLE AND MIREX

The application of audio gene recognition technology is very broad, and the application based on audio gene technology come true Music Information Retrieval by audio gene. And the basic technology can be divided into: Audio Beat Tracking, Audio Melody Extraction, Audio Chord Detection, Audio Onset Detection, Audio Drum

Detection, Audio Key Finding, Audio Tempo Detection, Multiple Fundamental Frequency, etc. [2]. The technology principle diagram of content identification based on characteristic value is shown below as figure 1.

FIG 1 CONTENT RECOGNITION TECH BASED ON SIGNIFICANT VALUE(DIGITAL FINGERPRINT)

The latest technology trends of Music Information Retrieval technology can be obtained in the International Society for Music Information Retrieval. And the advanced algorithms can be identified in the Music Information Retrieval Evaluation eXchange(MIREX) Evaluation institution. MIREX was hosted by ISMIRSEL, an international music information retrieval system of the University of Illinois at urbana-champaign (UIUC). With its impartiality and credibility, it has become an important event in the information retrieval field of music every year. It aims to provide a fair and credible evaluation platform for the various frontier technologies in the field of music information retrieval and music signal processing. The application technology of MIREX evaluation can be divided into: Audio Artist/Classical Composer Identification, Audio Cover Song Identification, Audio Mood Classification, Audio Similarity And Retrieval, Audio Tag Classification, Music Recommendation, Query By Singing/ream/Tapping, Symbolic Genre Classification, Symbolic Melodic Similarity, etc. Among them, the audio emotion retrieval classification and the audio similar retrieval occupied bigger proportion. ACRCLOUD won the champion in 2016 Mirex contest in the item of fingerprint recognition and humming, and its service customers include Alibaba, millet, CCTV net, Shanghai, Letv, meters. The audience's judgment to music's emotional classification is subjective, and the standard of music's emotional classification is directly related to the accuracy of audio gene technology identification [3-4]. About this, Mirex divides music emotions into five categories: excited, joyful, sad, witty, and angry. The emotional marks are used to mark music and audio, and some marks such as "happy" and "melancholy," can be clearly used to emotional recognition and can be applied directly to information retrieval systems. The results show that other signs, such as those associated with style and specialized instruments, can also be used for emotional identification. Emotional awareness can be seen as a kind of multiple level- tag classification or regression analysis problems, we try to use a set of audio information retrieval emotions to annotate each piece of music [5], the mood is expressed as a single multidimensional vector or a vector time series and let them spread over the meaning of an emotional space. That is, each dimension of a vector represents a single emotion or a pair of bipolar emotions [1, 6]. We can use different forms of data to estimate the emotional vector value of an audio or music with several styles. At the same time, auditory features are commonly used to predict the rate of pleasure and awake.

3 AUDIO GENETIC TECHNIQUES

This paper will discuss from the two aspects of the emotion-based audio classification retrieval and content-based audio gene technology. The emotion-based audio gene technology mainly used K planar piecewise regression algorithm, and the content-based audio gene technology research mainly includes audio beat notation, starting point detection, hum retrieval and beat tracking technology.

3.1 Research on Audio Gene Technology Based on Emotion

In order to improve the music emotion classification accuracy based on the regression, the K planar piecewise regression method was used [3] here. In the high- dimensional space made up of music characteristics and music emotions, we can directly solve nonlinear regression problems by hyperplane method found by multiple iterations, thus we can predict two-dimensional emotional variable values Valence and Arousal to classify audio or music emotion. In this process, a regression analysis task was created, and the emotion parameter regression analysis was applied to predict the music emotion [2]. This kind of emotional classification method applied automatic

classification labeling technology, which mainly includes the following technologies:

TABLE 1 INTRODUCTION OF SEVERAL TECHNOLOGIES

Technology	Technology content	Technology principle
TF-IDF Measure[3]	It is often used by text mining to retrieve simultaneous events within the mood sign and then form all kinds of phrases until there is no simultaneous event.	This simultaneous event retrieval method compares the specified threshold and specific value of the number of songs associated with the two signs and the minimum number of songs associated with each symbol.
Latent semantic analysis (LSA) [2-3]	A natural language processing technology, which used approximation method to approximate a language matrix into a low rank matrix.	In this case, this sparse matrix language matrix is used to describe the number of times each song is marked with a given label. For a database of thousands of songs and more than 100 possible mood symbols, the language matrix will be in very high order. After some operations, the deformed language file matrix is dealt with single value decomposition (SVD) to produce both left and right single vectors, which represents the distance between the terms and the file. Levy and Sandler at first used a variable called "correspondence analysis", and inferred a semantic space from a batch of Last.fm social marks collected, which covered more than 24000 specialized signs and 5700 tracks. Each vector is grouped by the calculated cosine distance between various signs, after that any random such as classification method is used to combine the language together.
K plane clustering algorithm	It is derived by the k-mean clustering algorithm, and the k-plane subsection regression method is derived.	The advantage is that the nonlinear problem can be solved directly and no need to select the kernel function and kernel parameters.

3.2 Research on Audio Gene Technology Based on Content

The technology principle of audio gene based on content is, analyze the lyrics as text messages during the process techniques of text mining to natural language processing are used [2]. Here set an integrated audio and lyrics comprehensive system for example, its working principle is, the audio and text features are integrated into a single vector for joint classification, improve operation efficiency in two quadrants exceed 5% pure audio features: "happiness" (81.5% to 86.8%) and "uncomfortable" (87.7% to 92.8%). Other quadrants remained largely unchanged, but the audio features were already highly classified: "leisure" (91.4% versus 91.7%) and "angry" (98.1 to 98.3%).

1) Mark the Music Method by the Beat

It is a kind of beat automatic tagging method based on the characteristics of CQT [7] and dynamic programming (DP) [8], mainly includes two kinds of algorithms: Const Q Transform and DP algorithm. One of the advantage of this method is to tag music beat very well for various style music, and the other is to overcome the label problem of not obvious beat [9-11].

TABLE 2 INTRODUCTION OF USED ALGORITHM

Algorithm	Algorithm principle	The advantage of algorithm
Const Q Transform	Different frequency resolution was obtained by calculated, thus to calculate frequency amplitude of each chromatic. according to the constant Q, by adopting different window width	It highlights the frequency characteristics of music notes, so it can better represent music signal

DP	Dynamic programming is a classical algorithm of divided the problem to solve into several sub-problems, and then calculate the solutions of sub-problems to obtain the optimal solution of the original problem	The current point integral calculation needs to refer to the first 3/2 to 1/2 of the time range, which is the integral of the data frames within a time frame to ensure that the last optimal node is within the search category

2) Study on the Starting Point of Music Notes

In the field of content-based music study, correctly extracting the notes starting point information is the basis to identify music advanced features such as pitch, rhythm, tempo, and paragraphs, thus the notes starting point detection based on matching pursuit algorithm [12] arises at the historic moment. Based on the interpretation degree [7,12] and the detection algorithm based on the change of the partial tone [13-14]are two new types of note starting point detection algorithms. Both of these algorithms analyze the code based on decomposition and use the improved peak extraction algorithm to generate the starting point vector of the notes. The advantages of these two kinds of algorithm firstly for the detection of mixed music has good effect, the other for the use of the improved peak extraction algorithm to generate notes starting point, thirdly to improve note point performance of the algorithm.

3) Music/Audio Humming Research

In order to improve the speed and accuracy of the melody matching, a matching algorithm based on frame-note mode [15] was proposed. Based on the shape of the melody curve, this algorithm uses the base frequency sequence to express the hum clip, so as to avoid the error caused by the sequence of notes. The note sequence is used to represent the template fragment, and then estimate base frequency jump point location according to the cumulative weight, then to calculate the edit distance between the humming fragment and the template fragment. The advantage of this algorithm is that it can optimize the matching process by using the phonon information and would not import new errors to the humming fragment. In addition, a music retrieval algorithm based on humming [16] was proposed to identify audio genes based on the humming music retrieval system. This algorithm innovatively constructed the melody histogram feature of music note sequence, which could express variable length note sequence as fixed 12 d melody histogram feature, thus import further the high-dimensional data index BVI (Bit Vector Indexing) algorithm into the humming retrieval systems, using melody histogram feature to set up BVI index to effectively and rapidly filter song candidate set [17]. This system applies gene frequency extraction method, music melody features based on the gene frequency extraction, note segmentation to extract music characteristics, then applies DTW (dynamic time warping) algorithm and the improved RA (Recursive Alignment) algorithm in order to improve the retrieval accuracy, embodied in changing distance measure, adding notes sequence and pitching contour matching endpoint penalty term and notes sequence offset.

The following is a description of the difference between RA algorithm and the improved RA algorithm in the music retrieval algorithm based on humming.

TABLE 3 THE CONTRAST BETWEEN RA ALGORITHM AND IMPROVED RA ALGORITHM

RA Algorithm	Improved RA Algorithm
- USES second - order distance to calculate distance - No need to require the information of the base frequency trajectory cutting points, so the processing of the base sound trajectory becomes to eliminate all zeros and set the average value as zero.	- In the distance calculation, the first order distance is used -The penalty item is added to the position of the base tone frequency shear point and the note endpoint mismatch, and the best match is attempted by the overall deviation of the sequence of the notes

4) Research on music beat tracking algorithm

Beat tracking is the detection to the "pulse" or the significant periodic music event[11]. In music information retrieval, the beat tracking is often used to chord identification, song detection, music segmentation and transcription, etc.[13] A

minimum and maximum-based distance algorithm[18] was proposed to extract music beats and music characteristics of any music type. Its core is to determinate the beat point starting, extract BPM significant value and effective peak value. It innovatively applied the clustering algorithm to beat tracking research, transformed the peak value extraction problem to classification problems and completed the beat extraction from the perspective of clustering sequence.

4 DOMESTIC AUDIO GENE TECHNOLOGY RESEARCH

Audio gene technology can be widely used in many fields, and as the trend of copyright protection prevails, more and more enterprises are engaged in related field research. Chinese enterprises have achieved excellent results in the evaluation of MIREX algorithm, such as the You Tu-team, which won the first and second place in the 2015 hum retrieval competition. In 2014, netqin music won the champion in the audio retrieval evaluation contest. In 2015, sogou speech recognition technology won three first and two second place in the five sub-items of the "hum technology" part. The following is a list of the companies and their research development to help musicians or companies to protect music or audio copyright.

4.1 Tencent YouTu Audio Recognition Technology

Tencent is the first group to start making artificial intelligence, focusing on face recognition, image recognition and audio recognition. Face recognition is mainly applied to the dynamic and static facial recognition system and living detection algorithm, in the service of the identity of the Banks and entrance guard system. The image recognition is mainly used in mobile phone QQ card recognition and work against eroticism. And the team operates voice recognition project of the figure as follows.

The humming recognition algorithm can automatically recognize a melody from the user and return the correct song to the user after the retrieval in the cloud, which aims to help the user to find the song whose name forgotten. This algorithm scored two first places and second places of the world in MIREX2015's humming competition, and the results has been published the in the top-level artificial intelligence conference AAAI2016.

The Acoustic recognition algorithm can collect played music by microphone and generate audio fingerprint, and then the collection would be sent to the cloud to retrieve and return the right song back. Now it can rely on the ten million level database of qq music to return real-time results in the library, and the accuracy is more than 98%.

With the technology of sound pattern recognition and deep learning technology, human can be recognized by voice, and it can reach a rate of 0.1% error acceptance rate and 95% pass rate

The automatic library scheme, based on the frontier audio technology and machine learning songs information completion and correction technology, can automatically complete or correct the wrong information, can automatically solve the problems including music classification not accurate, the lyrics time information error and the lack of foreign music lyrics subtitle, it can alternatives the traditional artificial selection plan, and promote the quality of library and the mass library construction efficiency and lower operating costs.

4.2 Net Qin Music Radar Audio Recognition Technology

In 2014, the winner of the international audio retrieval evaluation contest. Defeated the team of Columbia University and Peking University in the performance of the audio-music technology of (audio fingerprint) fingerprint competition, it has the background of cars automation institute. Net Qin team is now developing to smart cars, mobile entertainment, mobile games and mobile security.

4.3 Sogou Speech Recognition Technology

The Sogou speech recognition technology team won numerous awards in the 2015 international audio retrieval evaluation contest (MIREX), becoming the biggest winner in the competition. In the "humming technology" part, Sogou won three of the five sub-items, two second place, and the third place in the "listening to music" competition. The advantages of Sogou speech recognition technology are:

- Able to break through the mixed difficulties of China and the English,

- Sogou input method has a built-in language recognition engine
- In the wave of AI, the voice recognition modeling technology evolution is rapidly, and speech recognition model are also emerged in endlessly, especially language model includes the N - "gram, RNNLM etc, and the acoustic model covers the HMM, within DNN, RNN model

4.4 IFLY Voice Recognition Technology

The accuracy of speech recognition system has reached 97%, which has stood by Baidu and Google. The point that increase IFLY speech recognition accuracy is a new framework called feed-forward type sequence memory network FSMN, compared with the traditional model of deep learning , the feed-forward type sequence memory network adds a memory module to store the useful judged historical and future information for current frame, and establish a model called "end-to-end" in the speech recognition, to integrate a voice transcription engine combined with automatic speech segmentation, natural language processing and voiceprint recognition technology. The innovation is lots of custom work for the voice recognition in different scenes, such as the IFLY speech recognition system can realize the automatic separation of dialogue roles in meeting or interview.

To sum up, in recent years, China's Audio gene identification technology made a great progress even has achieved world-class level, especially in the field of Audio Mood Classification, Audio Similarity and Retrieval, Query By Singing/Humming, Tapping, China has been more stronger than other countries and China is playing a leading role now. In this paper, the application logic diagram of audio gene technology is established based on the technology of audio genes in China, as shown in figure 2. There are two kinds of audio gene technology users, one for audio humming retrieve and the other for users who need to upload audio. Audio humming retrieve users retrieve through the humming retrieval process, and generate the independent humming audio and then matched with the total audio gene pool, finally return the result to the users; For the users need to upload the audio, in the process of waiting for audit the audio would establish independent recognition file and generate the audio gene fingerprint after the audio gene detection. Then the fingerprint would be matched with the total audio gene pool, if the copy rate is lower than the given standard, the audio would be judged as low-repeat rate audio and would be added into the total pool as the normal operation, if not, the audio would be dealt as abnormal operation. And there are four main ways for the abnormal operation.

- Block: block the video uploaded or terminate the account
- Regional blockage: allows films containing infringing content to be viewed in specific areas
- Tracking: track the statistics of audience watching the infringing films and the most frequently viewed areas, which are the best marketing and publicity materials for copyright owners
- Profit: it can benefit the copyright owner, violators and audio publishing platform, copyright owner can insert ads in containing the infringing content in the film, and the profit produced by the process of viewers clicking on the ads would be shared by platform and copyright owner. What's more, other platforms will also provide the purchasing link to the infringing content, to allow the viewer to choose and buy related copyrighted music or video content. For the infringer, it is a disguised form of authorization
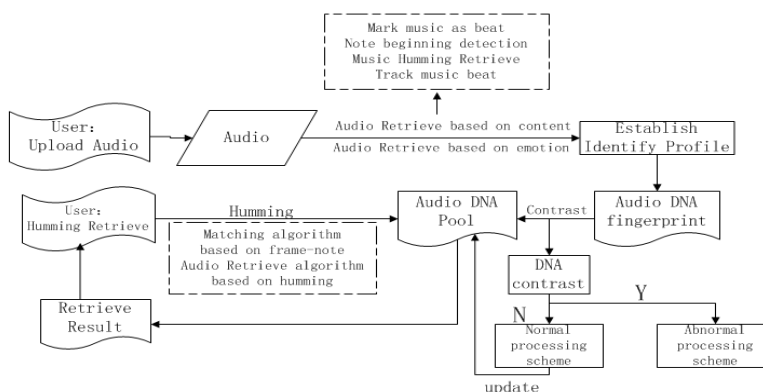
FIG 2 CHINA AUDIO GENE TECH APPLICATION LOGISIT DIAGRAM

5 CONCLUSIONS

Based on the analysis of the principle of audio gene technology and gene technology application direction, the following conclusions can be drawn:

- The domestic algorithm currently applied in audio gene technology is more focused on K plane clustering algorithm, CQT algorithm respectively by K, DP algorithm and DTW algorithm and some algorithms based on RA algorithm such as K planar piecewise regression algorithm,, note starting point detection algorithm, humming retrieval algorithm and the beat tracking algorithm, which complete the process of the beat, content and emotion from different views. And in the field of Audio Mood Classification, Audio Similarity and Retrieval, Query by Singing/I/I, etc., China has become the world leader.

- In this field, the leading domestic enterprises includes Tencent YouTu Lab ,Net Qin music team, sogou speech recognition and IFLY speech recognition systems, in the field of audio gene recognition technology they all have achieved high grades shown as the high accuracy of speech recognition, the method to overcome language difficulties, the perfect module function.

- However, compared with Google audio recognition technology, domestic enterprises in this field are lack of research about systematic rewards and punishments. The perfect rewards and punishment system of Google audio recognition technology must be supported by a large algorithm framework, and can ensure the smooth operation of database services. Thus we can say that, the domestic Audio gene Identification technology still need to strengthen and improve the algorithm framework build, and need to strengthen research about the Audio Artist/Classical Composer Identification, Audio Cover Song Identification, Audio Tag Classification, Music Recommendation, Symbolic Genre Classification and Symbolic Melodic Similarity technology such technology blank pot, to establish a more accurate, faster, more comprehensive Audio gene technology.

REFERENCES

[1] Sun Jian. Study on Real-time Rhythmic Information Retrieval of Musical Audio Signal and the System Implementation [D]. : Dalian University of Technology, 2011.

[2] Wang Huimin. Music Emotion Classification Based On Regression [D]. :Nanjing University of Posts and Telecommunications, 2015.

[3] Shao Xi，Wang Hui － min. Music Emotion Classification Based on k － plane Piecewise Ｒegression [J]. Computer Technology and Development, 2015, (6): 166-170.

[4] Ji Xi Cheng. A Study on Automatic Generation of Family Music Album Based on Emotion [D]. : Nanjing University of Posts and Telecommunications, 2016.

[5] Song Liming ,LI Ming,Yan Yonghong. Pitch estimation based on harmonic salience [J]. Acta Acuatica, 2015, (2): 294-299.

[6] Xu Jie-ping，YIN Hong-yu，FAN Zi-wen. Study on cover songs identification based on phrase content [J]. Journal of Shandong University(Natural Science, 2013, (7): 68-71, 78.

[7] Gui Wen-ming, LIU Rui-fan, SHAO Xi, BAI Guang-yi. Note Onset Detection Based on Constant Q Transform [J]. Computer Engineering, 2013, (10): 283-286.

[8] Zhang Wei-wei，CHEN Zhe，YIN Fu-liang，ZHANG Jun-xing. Ｒeview on Melody Extraction from Polyphonic Music [J]. Actaelect Ronica Sinica, 2017, (4): 1000-1011.

[9] Wang Yue,Xie Lei,Yang Yulian. Adaptive whitening for real-time music beat tracking [J]. Application Research of Computers, 2009, (5): 1676-1678, 1684.

[10] Wang Tianjiang ,Chen Gang. A new method of Beat Tracking Based on CQT Features [C]//（HHME2007.）

[11] Li Na. Research and Application of　Real-Time Music Beat Tracking System [D]. : Shandong University, 2011.

[12] Actaelect Ronica Sinica, 2013, (6): 1225-1230.

[13] He Ning. Beat Tracking Based on Max-Min Distance Means [D]. :Tianjin University, 2014.

[14] Wang Tianjiang ,Chen Gang. A new method of Beat Tracking Based on CQT Features [C]//（HHME2007.）: 107-113.

[15] Guo Min, Zhang Weiqiang, LIU Jia. A frame-to-note algorithm for queryby humming [J]. J Tsinghua Univ(Sci& Tech),, 2011, (4): 561-565.

[16] Li Cong. Query By Singing and Humming system [D]. :Tsinghua Univ, 2009.

[17] A J Ghias,Logan D Chamberlain.Qurey by humming musicalinformation retrieval in an audio database..SanFrancisco.

[18] He Ning. Beat Tracking Based on Max-Min Distance Means [D]. : Tianjin University, 2014.

AUTHORS

[1]Cui Jizhe is a professor of Dept. of Information Management and Systems at YANBIAN University in China. He received the Undergraduate course Degree from Northeast Normal University, Major in Mathematics Education in 1996. Received the Ph.D. degree from Sangmyung University, major in Computer Science in 2008. His research interests are digital watermarking, copyright management, Data analysis and decision making, Data Modeling, and BigData Analysis.E-mail:cuijizhe@foxmail.com.

[2]Tan Xinjie, graduate of technical economics and management in Yanbian University. Major in patent collection and analysis, and has been published 2paper in the field.E-mail:xjenactus@foxmail.com.

·