法政大学学術機関リポジトリ

HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-07-04

Singer Identification of Pop Music with Singing-voice Separation by RPCA

Lu, Xing

(出版者 / Publisher)
法政大学大学院情報科学研究科

(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編

(巻 / Volume)
13

(開始ページ / Start Page)
1

(終了ページ / End Page)
6

(発行年 / Year)
2017-03-31

(URL)
https://doi.org/10.15002/00021518

Singer Identification of Pop Music with Singing-voice Separation by RPCA

Lu Xing
Graduate School of Computer and Information Sciences
Hosei University
Tokyo 184-0003, Japan
Lu.xing.99@stu.hosei.ac.jp

***Abstract*—Singer identification is important in music organizing and retrieving. However, in many cases, the correct rate of singer identification system is not high enough. In this paper, we propose an effective system of singer identification with human voice separated from original music. The first part of this research is music separation, and we would like to use Robust principal component analysis (RPCA) to solve this problem with its high performance. After the clear enough human voices are extracted, we can proceed to the second part, singer identification. At this stage, the Linear Predictive Coding (LPC) method was chosen as the experimental method. When we finish extracting the LPC features, the singer would be identified by Gaussian Mixture Model (GMM). MATLAB is used to the singer identification algorithm.**

Keywords—Music Separation; RPCA; Singer Identification; GMM

I. INTRODUCTION

These years, with the rapid development of digital music, singer identification and lyric alignment has attracted more and more concern. Music separation which is an important part of these techniques has been valued [1]. With the pursuit of spiritual satisfaction, music spread more and more widely, and people are exposed to music every day [2]. Nowadays, A song generally consists of combination of vocal portion and various musical instruments [3]. (human voice, guitar, bass, drum etc.), maybe it is quite easy for human beings to recognize a target of audio, but it is hard for a machine to do so [4]. Sometimes people hear some songs and like them a lot, but they do not know any information about the songs at all. Therefore, technologies are demanded for efficient categorization and retrieval of these music collections, so that consumers can be provided with powerful functions for browsing and searching musical content.

With this capability provided in a music system, the user can easily get to know the singer's information of an arbitrary song. Currently, singer's information is manually embedded into music files by professional recorders. However, such information is often lacking, or inconsistent in music pieces downloaded from music-exchanging websites, or music clips grasped from digital music channels. Singer identification is a technique to mimic human’s ability to know who is singing in the currently playing music [5].

There are previous technologies on the automatic speaker identification which identify the speaker of a given speech segment. While the basic idea of the singer identification is also to recognize the voice of a human being, there are significant differences between a singing signal and a speech signal in several aspects. First of all, the singing voice is mixed with musical instrumental sounds in a song, which makes it much more complicated to extract features of the voice. Furthermore, the time-frequency features of a singing voice are quite different from those in a speaking voice. Therefore, we believe that by extracting and analyzing audio features properly, an automatic system should be able to achieve certain degree of singer identification as well.

II. RELATIED WORK

Up to now, there are many music separation techniques, such as non-negative matrix factorization(NMF) [6], robust principal component analysis (RPCA) and low-rank decomposition. NMF iterative process, is completely unsupervised, no information is given about accompaniment and vocals. So, the NMF algorithm is not suitable for music separation whose musical instrument source information is unknown. But if you know the music composition of musical instruments, the effect of NMF algorithm will be improved. The separation performance of RPCA is better than that based on non-negative matrix factorization, and it takes less time, easy to implement. A supervised low-rank decomposition algorithm can effectively separate the accompaniment and vocals from the music, but training these dictionaries require a lot of time and a lot of corpus. When enough dictionaries are lacking, the separation of the low-rank decomposition algorithm is reduced [7]. So, there is a need to strike a balance between computational complexity and good separation.

These algorithms have their own advantages and disadvantages, but the common drawback is that the vocal and accompaniment cannot be completely separated.

There are similarities between singer identification and speaker recognition, but the inherent difficulties lie in the nature of the problem: the voice is usually accompanied by

Supervisor: Professor Katunobu Itou

other musical instruments and even though humans are extremely skillful in recognizing sounds in acoustic mixtures, interfering sounds usually make the automatic recognition very difficult [8].

The study of speaker identification dates back to the 1930s. Early work mainly focused on the possibility of perception experiment of human ear and exploring listening recognition. Since the 1960s, the study of speaker recognition has focused on the linear or non-linear processing of various acoustic parameters and the new pattern matching method, such as dynamic time regularization, principal component analysis, hidden Markov model, neural network and multi-feature combination technology [9]. So far, speaker recognition has developed increasingly, and applied in some projects successfully.

But there is still a lot of difficulty in the speaker recognition, it’s still unsatisfactory until today [10]. This is mainly caused by the speaker's feature extraction problem, which is attributed to the following aspects:

A. **Simple and reliable speaker voice feature parameters**: The speech signal contains both the semantic information of the speech content and the personality information of the speaker's vocal characteristics. It is a mixture of the phonetic features and the speaker's characteristics. So far, there is no good way to separate the speaker's individual characteristics from the phonetic features, nor to find simple acoustic parameters that can reliably identify the speaker.

B. **The variability of speech signals**: Even for the same speaker and the same text, the voice signal is also very different. The speaker's speech feature is not static, it has time-varying characteristics. The variability of the speech signal moves the speaker's character space essentially and the speaker pattern changes, thereby increasing the uncertainty in the recognition process.

In addition, the application of the singer recognition is also troubled by problems such as unclear vocals. Therefore, when we extract the voice features, we only extract the vowels which most people pronounce more standard.

III. PROPOSED APPROACH

A. Overview of the proposed approach

In order to achieve the experimental purpose of the singer identification, I proposed an accurate and efficient singer recognition system, with RPCA to separate music, LPC to extract the vocal features and GMM to identify singers. The system procedure is described as Fig. 1.

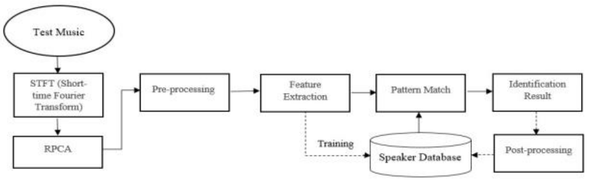


Fig. 1. Procedure of Proposed System

B. Basic Flow of Singer Identification System

Our objective is to introduce an effective system for singer identification that minimizes all aspects of the error. During the research preparation phase, ten well-known singers were selected as samples, including five male singers/group (Only one person's voice is in the music) and five female singers. Then we randomly selected ten or more songs from each singer's music. After the acoustic material is ready, we need to find the most suitable music separation method for this research. Three alternatives were found by searching the literature: Non-negative matrix factorization(NMF), Robust principal component analysis(RPCA), Low-rank decomposition. We chose RPCA based on time-frequency decomposition, because the algorithm separates the music more thoroughly and has strong robustness, and the overall performance is the best among the three algorithms.

In this research, RPCA was used to separate music. The goal of RPCA is to decompose a matrix into a sparse matrix and a low rank matrix, under the condition that L + S = X is satisfied, the following formula is minimized:

$$\|L\|_* + \lambda\|S\|_1$$

Where X is the data matrix of the mixed signal, S and L are the decomposed sparse and low rank matrices, respectively, and $X \in R^{n1xn2}$, $L \in R^{n1xn2}$, $S \in R^{n1xn2}$.

The above X is the mixed music signal, S and L correspond to the background accompaniment and the human voice signal part, the algorithm is as follows:

1) *The matrix X is obtained by calculating the spectrum of the mixed music signal using the Short-time Fourier Transform (STFT).*
2) *we use the inexact Augmented Lagrange Multiplier (ALM) method [11], which is an efficient algorithm for solving RPCA problem, to solve L + S = |X|, given the input magnitude of X.*
3) *We can obtain two output matrices L and S: sparse matrix S, which indicates vocal activity, and musical notes in the low-rank matrix L.*
4) *In order to recover the time-domain waveforms of the vocals and accompaniment sections from the estimated signal, it is necessary to record the original signal phase P = phase (X), and the phase information is added to the matrices S and L using L (m, n) = $Le^{jP(m, n)}$ and S (m, n) = $Se^{jP(m, n)}$, where m = 1, …, n_1, n = 1, …, n_2. Finally, ISTFT can be calculated to obtain the time domain signal [12]*

During separating the music, we want to find another method to separate music to evaluate the performance of RPCA, eventually we chose another music separation algorithm called REPET-SIM [13]. It is easier to understand than RPCA, and the actual operation is also very simple. But after testing this algorithm, the results show that this algorithm isn't suitable for all songs. For example, while we separated a song called Hold It Against Me, with masking hardness and masking threshold were 0.8 and 0.4, this algorithm performed very well. However, for another song called Counting Stars, although the parameters were adjusted to the most appropriate situation, the song's vocal and accompaniment is still hardly separated. Therefore, REPET-SIM is not reliable for this research.

The basic concept of Linear Predictive Coding (LPC) is that a speech signal can be approximated by the linear combination of several previous speech samples, by minimizing the sum of the squares of the difference between the input speech samples and the linear prediction sample values, a unique set of prediction coefficients can be determined.

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

In the LPC model, the linear time-varying system of the representative channel model is estimated, and actually the parameter a_k of the system function H(z) is estimated. Assuming that the input speech is short and stable, that is, in a short period of time, it can be considered that the excitation source and channel model do not change, a_k varies only with each short period of time. Obviously choosing a frame of voice signals to determine the a_k, and we get the channel model parameters which describe this short time. The channel model parameter represents the channel characteristics of different people at a given time. The resulting parameter a_k is called the LPC characteristic parameter.

The LPC characteristic parameter embodies the channel resonance characteristic of the speech signal and is widely used in the speech signal processing. The solution can be applied by the autocorrelation method. Proceed as follows:

1. The input speech signal s(n) is windowed to obtain the short-term speech signal s'(m).

   $$s'(m) = s(n+m)w(m)$$

   Where w (m) is a window function (m = 1, 2, ..., N) and N is a frame length.

2. Calculate the short-term autocorrelation function R (k) of s' (m)

   $$R(k) = \sum_{m=0}^{N-1-k} s'(m)s'(m+k) \qquad k=0,1,\ldots,p$$

3. a_k is the linear prediction coefficient, k = 0,1,2, ..., p, the following matrix equation is obtained.

   $$\begin{bmatrix} R(0) & \cdots & \cdots & R(p-1) \\ R(1) & \cdots & \cdots & R(p-2) \\ \cdots & \cdots & \cdots & \cdots \\ R(p-1) & \cdots & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix}$$

   a_k, k = 0,1,2, ..., p can be obtained with the above equation, and the autocorrelation matrix is a Toeplitz matrix.

The Gaussian mixture model is to estimate the probability density distribution of the sample, the estimated model is the sum of several Gaussian models weighted. Each Gaussian model represents a cluster, the data in the sample are projected on several Gaussian models, respectively, and the probability of each cluster is obtained. The Gaussian mixture model is shown in Fig. 2, which is obtained by weighted summing of M multidimensional Gaussian distributions.

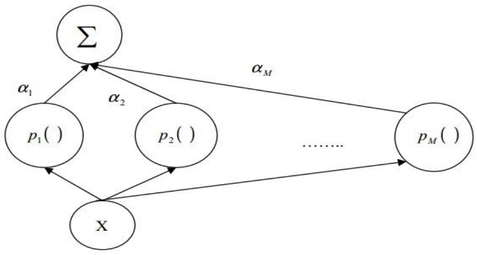


Fig. 2. Gaussian Mixture Model Schematic

It can be expressed in mathematics as:

$$p(x_t) = \sum_{i=1}^{M} \alpha_i p_i(x_t)$$

x_t is the D-dimensional speech feature vector; p_i (x_t) is the Gaussian mixture model component, it is the D-dimensional Gaussian distribution function; α_i is the weighting coefficient of the corresponding component p_i (x_t); M is the number of components in the Gaussian mixture model. For p_i (x_t) and α_i, they satisfy the following equation:

$$p_i(x_t) = \frac{1}{(2\pi)^{\frac{D}{2}} \left|\sum_i\right|^{\frac{1}{2}}} \exp\left\{ -\frac{(x_t - \mu_i)^T \sum_i^{-1} (x_t - \mu_i)}{2} \right\}$$

$$\sum_{i=1}^{M} \alpha_i = 1$$

It can be seen that the respective components p_i (x_t) of the Gaussian mixture model can be described by the mean vector μ_i and the covariance matrix Σ_i. Therefore, GMM can be expressed by the parameter set $\lambda = \{\alpha_i, \mu_i, \Sigma_i$ (i = 1, 2, ..., M)}.

In GMM, the personality characteristics of each speaker are uniquely determined by the Gaussian mixture probability density function of different parameter values. Therefore, in the training process, the system not only to estimate the specific speaker corresponding parameter λ, but also to get the speaker speech feature sequence probability of the largest parameter λ. Expectation-Maximization(EM) algorithms are often used to estimate Gaussian mixture model parameters.

IV. IMPLEMENTATION

In this part, we will introduce the detail of implementation.

A. Music Separation

First of all, the target audio signal is processed to time-frequency analysis by using a Short-time Fourier Transform. The Short-time Fourier Transform (STFT) is a digital processing in which the time-domain data of the digital samples are decomposed into chunks that are usually overlapped. And Fourier transform is used to calculate the magnitude of the frequency spectrum of each chunk. The sinusoidal frequency and phase content of the signal could be determined by this method.

The Augmented Lagrange multiplier method is used to solve the above RPCA problem with superior convergence. The ALM algorithm is basically an iterative convergence scheme that works simultaneously by minimizing the rank of the sparse matrix and low-rank matrix repeatedly.

Inverse Short-time Fourier Transform is used to recover the audio signal in time domain and the results are evaluated.

B. Singer Identification

In general, the spectrum of the speech signal has a high-frequency attenuation phenomenon. Normally, a very simple first-order FIR filter is used to perform pre-emphasis before doing LPC analysis.

For the linear predictive coding, we use MATLAB to implement it. First let it pass through the FIR filter, then create a 1024-point Hanning window for the audio signal. And we set the order of the prediction filter polynomial to 21. In the case of these parameters, we continued LPC analysis.

In training GMM, the K-means algorithm is used to obtain the initial value of the model parameters, and then the maximum expectation algorithm is used to estimate the model parameters. The probability density function of GMM is as follows

$$p(X|\lambda) = \sum_{i=1}^{M} \alpha_i p_i(X|\lambda_i)$$

P (X | λ) is obtained by mixing the M Gaussian mixture model according to the mixing coefficient α_i.

The logarithmic likelihood function of the nonholonomic sample set X is expressed as

$$\log(L(\lambda | X)) = \log\prod_{t=1}^{T} p(x_t | \lambda) = \sum_{t=1}^{T}\log(\sum_{j=1}^{M}\alpha_j p_j(x_t | \lambda_j))$$

Where T is the number of samples in X and M is the number of Gaussian mixture models in the mixed model. Considering that X is a non-complete case, assuming that there is an unobserved data item Y ≈ $\{yt\}^{T}_{t=1}$, the value of Y indicates that the sample in each X set is generated by which individual model of the mixed model. After inputting Y, the likelihood function of the complete sample set can be written as

$$\log(L(\lambda | X,Y)) = \log(p(X,Y | \lambda)) = \sum_{t=1}^{T}\log(\alpha_{y_t} p_{y_t}(x_t | \lambda_{y_u}))$$

For the Gaussian mixture model, the mathematical expectation of the likelihood function of the above equation is calculated according to the E step of the EM algorithm

$$Q(\lambda,\lambda^E) = E[\log(L(\theta|X,Y))]$$

The above formula can be expanded to

$$Q(\lambda,\lambda^E) = \sum_{l=1}^{M}\sum_{t=1}^{T}\log(\alpha_l)p(l|x_t,\lambda^E) + \sum_{l=1}^{M}\sum_{t=1}^{T}\log(p_l(x_t|\lambda_l))p(l|x_t,\lambda^E)$$

Where l ∈ 1,2 ... M, p (l | x_t, λ^E) represents the posterior probability of a sample x_t belonging to the Lth model in the mixed model in the case of the sample set X and the model parameter λ^E are known. Calculated by the following formula, k represents the kth cycle

$$p^{(k)}(l | x_t,\lambda^E) = \frac{\alpha_l^{(k)} p_l(x_t | \lambda_l^{(k)})}{\sum_{j=1}^{M} C_j^{(k)} p_j(x_t | \lambda_j^{(k)})}$$

In the step M of the EM algorithm, we need to find the model parameters that maximize the likelihood function. It can be seen from the expansion of the above likelihood function Q (λ, λ^E), which is obtained by adding two items. The former is related only to C_l, and the latter is related only to λ_l. Maximizing the former term yields a new mixed coefficient estimate, maximizing the latter to obtain a new mean and covariance matrix estimate.

The probability that the feature vector falls into the hidden state i is

$$p(i | x_t,\lambda) = \frac{\alpha_i p_i(x_t)}{\sum_{kl=1}^{M}\alpha_k p_k(x_t)}$$

Thus, you can get the following three GMM parameter revaluation formula. The weighting coefficient revaluation formula is

$$\alpha_i^{(k+1)} = \frac{1}{T}\sum_{t=1}^{T} p^{(k)}(i | x_t,\lambda)$$

The mean vector revaluation formula is

$$\mu_i^{(k+1)} = \frac{\sum_{t=1}^{T} p^{(k)}(i | x_t,\lambda)x_t}{\sum_{t=1}^{T} p^{(k)}(i | x_t,\lambda)}$$

The variance revaluation formula is

$$\sigma_{ik}^2 = \frac{\sum_{t=1}^{T} p^{(k)}(i \mid x_t, \lambda)(x_{tk} - \mu_{ik}{}^{(k+!)})^2}{\sum_{t=1}^{T} p^{(k)}(i \mid x_t, \lambda)}$$

The iterative repetition of the above formula is the repetition iteration of step E and step M of the EM algorithm. Stop the iteration when the maximum of the likelihood function is found.

Gaussian mixture model is also implemented in MATLAB, first of all, we set N and k to 150 and 5 respectively. Then we allowed for uniform background noise term, and we set the number of repetitions with different initial conditions to 10, but only the best fit is returned. In the EM algorithm, maximum iteration number was set to 100, and tolerance value is 0.01. In addition, This MATLAB function requires the Statistics and Machine Learning Toolbox.

However, in the GMM training phase identified by the speaker, due to possible training data and other factors, it causes the EM algorithm to appear singular matrices, which is a significant defect. According to the characteristics that singular matrix does not appear in maximum likelihood estimation, the maximum likelihood estimation model can be used as the initial iterative model. Then in the iterative process of each step in the EM algorithm, we can use the given parameter α to control the correction ratio to modify it.

V. RESULTS EVALUATION

After using RPCA to separate music, we get background accompaniment and vocals, Fig.3 shows the spectrogram of original music, human vocal and accompaniment.

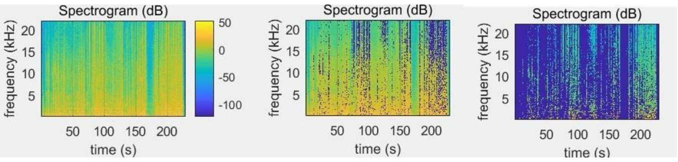
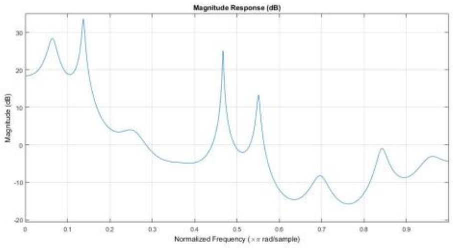


Fig. 3. Spectrogram of Music Separation

In the ten selected singers, pick out those songs whose human voice are clear enough. Find clear and noiseless vowel “o” in these vocals. Using LPC analysis for the same vowel made by these different singers, and some examples are shown in Fig. 4.



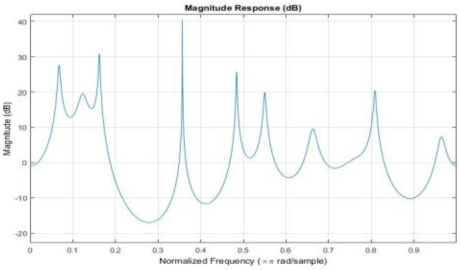


Fig. 4. Male and Female Singer LPC Analysis

From Fig.4, we can easily see the two singers different sound characteristics, but by virtue of this, the machine cannot tell the attribution of different sounds. The main reason for using the Gaussian mixture model in the singers' identification is that it can smoothen the density of any shape smoothly and achieve the closest result to the original data distribution. And the more the order of mixing, the more it can achieve the purpose of approximating the original data distribution, but the required identification time will increase. GMM experimental conditions are as follows: ten singers, half male, and the other half are female. In order to be more realistic, we picked out 16 songs for each singer (an album), five for training, and the rest for testing. each person's voice training time is about 5 seconds, test voice unit length is 1 second, using 21-order LPC as the basic characteristics, and the order of GMM is 5. Finally, we can use the following formula to calculate the recognition accuracy, and the accuracy is 81.8%.

$$\text{Correct recognition rate}(\%) = \frac{\mathit{Total\ number\ of\ correctly\ identified\ recordings}}{\mathit{Total\ number\ of\ testing\ recordings}} \times 100\%$$

Table.1. Confusion Matrix of The Identification(RPCA)

Singer number	1	2	3	4	5	6	7	8	9	10	Accuracy (%)
1	**8**	0	0	1	0	0	0	1	1	0	72.7
2	1	**9**	0	0	0	1	0	0	0	0	81.8
3	0	0	**7**	0	0	2	1	0	0	1	63.6
4	1	0	0	**10**	0	0	0	0	0	0	90.9
5	0	0	0	0	**10**	0	0	1	0	0	90.9
6	0	1	1	0	0	**9**	0	0	0	0	81.8
7	0	0	0	0	1	0	**10**	0	0	0	90.9
8	0	0	0	0	2	0	0	**8**	0	1	72.7
9	0	0	0	0	0	0	0	0	**11**	0	100
10	1	0	1	0	0	0	0	1	0	**8**	72.7

In Table 1 can be seen, the singer #9 got the best identification result, reached 100%, and the results of singer #4, #5 and #7 which are more than 90% are also good enough, while singer #3 only succeeded in identifying seven songs in eleven. On the whole, more than half of the singers' identification results are more than 80%, which shows that this method can be used as a singer identification method.

This paper described a method that performs singing voice separation and singer identification. The implementation of GMM in this system is based on the existing signal processing algorithm, but I found out that the order of GMM is also one of the factors that influence the accuracy of identification. At the beginning of the experiment, the reason I chose the GMM order to be 5 was to avoid the complicated calculation caused by the large order. When the order is low, each Gaussian component obtained by training is smoothed by the fact that it contains too much difference, and the identification performance is poor. However, when the order is large, since each Gaussian component is derived from very few data, it becomes very sharp, which reduces the robustness of the training model.

VI. Discussion

This paper described a method that performs singing voice separation and singer identification. In order to evaluate the performance of RPCA in the separation of music on the singer identification, we randomly selected a male singer and a female singer from ten. Then, we used REPET-SIM to separate the 32 songs (2 singers×16 recordings). We measured separation performance by Source-to-Distortion Ratio (SDR), Sources-to-Interferences Ratio (SIR), and Sources-to-Artifacts Ratio (SAR). Higher values of SDR, SIR, and SAR suggest better separation performance. We chose to use those measures because they are widely known and used, and also because they have been shown to be well correlated with human assessments of signal quality. The value of SDR, SIR and SAR are shown in the following table.

Table.2. Comparison of separation performance parameters

	SDR	SIR	SAR
REPET	3.2	10.2	5.6
RPCA	6.3	12.7	7.6

The results could easily show that the performance of REPET-SIM cannot reach the same level as RPCA. For example, SAR means the ratio of signal energy to the error due to artifacts, the expression is $SAR = 10\log_{10}\left(\frac{\|s_{target} + e_{interf}\|^2}{\|e_{artif}\|^2}\right)$, where s_{target} is an allowed distortion of source, e_{interf} and e_{artif} represent respectively the interferences of the unwanted sources and the artifacts introduced by the separation algorithm. From the expression, we can draw that when separating the same music, the smaller the value of the artifacts introduced by the separation algorithm, the greater the value of SAR. Thus, RPCA introduces fewer artifacts, which means RPCA is a better music separation method. And the average identification accuracy of the two singers only reached 63.6% (male singer: 54.5%; female singer: 72.7%).

Table.3. Confusion Matrix of The Identification(REPET-SIM)

Singer number	1	2	Accuracy (%)
1	6	5	54.5
2	3	8	72.7

In summary, the performance of robust principal component analysis as a music separation method is better than REPET-SIM, it is more suitable for singer identification.

VII. Conclusion and future work

In this paper, we proposed a method of using human vocals separated from pop music to identify singers. First, we selected 10 singers (five male and five female) as samples, and picked 10 songs for each singer. Then, those music were separated by robust principal component analysis. And using LPC to analyze the vowels in those human vocals, we chose to analyze vowels, because the LPC analysis is based on the assumption of the all-pole model, it cannot accurately describe unvoiced sound and nasal sound. And finally use GMM to identify them.

As can be seen from the results, this system can work well. However, due to limited time and effort, the accuracy is not ideal. In the future, we will try some improved programs, such as trying other vocal analysis methods and parameter changing, or using more training data and test data to get more accurate results.

References

[1] Huang P S, Chen S D, Smaragdis P, et al. Singing-voice separation from monaural recordings using robust principal component analysis[C]. Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012: 57-60.

[2] Zhang, Tong. "Automatic singer identification." Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on. Vol. 1. IEEE, 2003.

[3] Burute H, Mane P B. Separation of singing voice from music accompaniment using matrix factorization method[C]. Applied and Theoretical Computing and Communication Technology (iCATccT), 2015 International Conference on. IEEE, 2015: 166-171.

[4] Masood, Sarfaraz, Jeevan Singh Nayal, and Ravi Kumar Jain. "Singer identification in Indian Hindi songs using MFCC and spectral features." Power Electronics, Intelligent Control and Energy Systems (ICPEICES), IEEE International Conference on. IEEE, 2016.

[5] Chang P. Pitch oriented automatic singer identification in pop music[C]. Semantic Computing, 2009. ICSC'09. IEEE International Conference on. IEEE, 2009: 161-166.I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[6] Souviraà-Labastie, Nathan, Emmanuel Vincent, and Frédéric Bimbot. "Music separation guided by cover tracks: Designing the joint NMF model." Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. IEEE, 2015.

[7] Chan, Tak-Shing T., and Yi-Hsuan Yang. "Informed Group-Sparse Representation for Singing Voice Separation." IEEE Signal Processing Letters 24.2 (2017): 156-160.

[8] Mesaros A, Virtanen T, Klapuri A. Singer Identification in Polyphonic Music Using Vocal Separation and Pattern Recognition Methods[C]. ISMIR. 2007: 375-378.

[9] Alku, Paavo, and Rahim Saeidi. "The Linear Predictive Modeling of Speech from Higher-Lag Autocorrelation Coefficients Applied to Noise-Robust Speaker Recognition." IEEE/ACM Transactions on Audio, Speech, and Language Processing (2017).

[10] Bhattarai, Kritagya, et al. "Experiments on the MFCC application in speaker recognition using Matlab." Information Science and Technology (ICIST), 2017 Seventh International Conference on. IEEE, 2017

[11] Lin, Zhouchen, Minming Chen, and Yi Ma. "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices." arXiv preprint arXiv:1009.5055 (2010).

[12] Umap, Priyanka K., Kirti B. Chaudhari, and Madhuri A. Joshi. "Unsupervised singing voice separation from music accompaniment using robust principal componenet analysis." Industrial Instrumentation and Control (ICIC), 2015 International Conference on. IEEE, 2015

[13] Sharma, Shivam, and V. K. Mittal. "Window selection for accurate music source separation using REPET." Signal Processing and Integrated Networks (SPIN), 2016 3rd International Conference on. IEEE, 2016.