SUPERVISED LEARNING APPROACH FOR SINGER IDENTIFICATION IN SRI LANKAN MUSIC

Rajitha A. Amarasinghe, Lakshman Jayaratne

University of Colombo School of Computing, Sri Lanka.

rajithaamarasinghe@hormail.com, klj@ucsc.cmb.ac.lk

ABSTRACT: *This paper describes a method of modeling the characteristics of a singing voice from polyphonic audio signals, in the context of Sri Lankan Music. The main problem in modeling the characteristics of a singing voice is the negative influences caused by accompaniment sounds. Hence the proposed method consists of a procedure to reduce the effect of accompaniment sound. It extracts the predominant melody frequencies of the music file and then resynthesize it. Melody is extracted only on the vocal-parts of the music file to achieve better accuracy. Features vectors are then extracted from the predominant melody frequencies, which are then subjected to supervised leaning approaches with different classifier, using Principal Component Analysis as feature selection algorithms. The models trained with 10-fold cross validation and different combinations of experiments are done to critically analyze the performance of the proposed method for Singer Identification.*

KEYWORDS***: Music Information Retrival, Singer Identification, Singing Voice, Voice, Vocal Timbre Similarity***

INTRODUCTION

Music Information Retrieval is a growing research area with many milestones been achieved throughout the researches carried out in the world. It's been helpful in many ways to identify, classify, process, authenticate music, generating transcripts and many more. The singing voice is known to be the oldest musical instrument that most people have by nature and plays an important role in many musical genres, especially in popular music. As the singing voice is important, the representation of its characteristics is useful for music information retrieval (MIR). For example, if the name of a singer can be identified without any information of the metadata of songs, users can find songs sung by a certain singer using a description of singers' names

Identifying the singer of a musical piece from the audio signals without any use of the information provided by the metadata of the particular musical piece and identifying the percussion and non-percussion instruments is also a research area growing where many would benefit of the results. Automatic identification of the singer of a musical piece, classification of music by its author, defining singing patterns of a singer would be many future possibilities of researches based on singer identification, while instrument identification of polyphonic audio signals would contribute towards auto generation of musical transcript, chord identification and so on.

What is interesting about this research problem is the level of difficulty in identifying these specific sounds or voices with the accompaniment of other musical instruments which comes as single polyphonic audio signal. What this research will focus on is to improve researches done in this regard and find novel ways in which we can effectively address the research problems.

LITERATURE REVIEW

Many researches has been conducted in the field of Musical Information Retrieval (MIR), which is a broader area of study. Out of many such research papers related to MIR I found a few researches done on related to instrument identification, singer identification and sound source separation. Efficient and intelligent music information retrieval (MIR) is a need of the 21st century. MIR addresses the problem of querying and retrieving certain types of music from large music data set. A singing voice is one of the key elements of music. As most part of music is characterized by the performing singer, analysis of singing voice reveals many characteristics of a song. The unique qualities of a singer's voice make it relatively easy for carrying out numerous tasks in MIR. The singing voice is completely characterized by its acoustic features. Acoustic features like timbre, vibrato, pitch and harmony describe the singing voice in the music and these are discussed in the paper. There are many applications of MIR which considers overall features of music, but the paper presents a review of those applications of MIR concerned directly to singing voice in the music. Also the paper lists the feature extraction methods and identifies the suitable feature appropriate for individual task of MIR.

As a major product for entertainment, there is a huge amount of digital musical content produced, broadcasted, distributed and exchanged. This growing demand of amount of music exchange using the internet, and the simultaneous interest of the music industry to find proper means to deal with the new way of distribution, has motivated research activity in the field of MIR [1]. There is a rising demand for music search services. Technologies are demanding for efficient categorization and retrieval of these music collections, so that consumers can be provided with powerful functions for browsing and searching musical content [2]. A singing voice is one of the key elements of music. Singing voice is one of the less studied vocal expressions and analysing singing voice is an interesting challenge. Singing voice differs from every day speech in its intensity and dynamic range, voiced duration (95% in singing voice whereas 60%in speech), formant structures and pitch. Moreover the singing voice has loud, non-stationary background music signal which makes its analysis relatively more complex than speech [3], [4]. Thus speech processing techniques that were designed for general speech are not always suitable for the singing voice. But major of the earlier work have tried to extend speech processing to the problem of analysing music signals.

MIR performs the task of classiûcation in which it assigns labels to each song based on genre, mood, artists, etc. Those tasks directly related to song classiûcation analysing the singing voice are Singer Identification, Singer Verification, Music annotation etc. Other extended applications involve distinguishing between trained and untrained singer, analyse vocal quality, vocal enhancement etc. The paper provides an overview of features and techniques used for the above classiûcation tasks. It provides a summary of different applications based on singing voice and maps the application to its best suitable acoustic feature, the extraction method of that feature and the appropriate classifier. The performance parameters that are essential to evaluate the system are also presented.

Sound source separation is vital in identifying the singer and instruments in a polyphonic audio signal. Out of the literature written about source separation, Fujihara et al. [1], identifies two methods on the accompaniment sound reduction and reliable frame selection for source separation which will help in terms of separating the voice and identifying different instruments played in the music piece. Further literature by Heittola et al. [2] suggests a source-filter model using non-negative matrix factorization for source separation in polyphonic audio signals. They seem to find difficulties in identifying some instruments and some notes played by the instruments. There is a clear problem of identifying music notation of instruments and it becomes more difficult when the time periods of a note is lower and lower.

In the literature by Virtanen [3], an unsupervised learning algorithm for the separation of sound sources in one-channel music signals is presented. The algorithm is based on factorizing the magnitude spectrogram of an input signal into a sum of components, each of which has a fixed magnitude spectrum and a time-varying gain. Each sound source, in turn, is modelled as a sum of one or more components. There is also a comparison of the proposed model with the independent subspace analysis and basic nonnegative matrix factorization, which could be a helpful input in my research.

Under singer identification, several researches has been done and they have displayed substantial results. Among the literature written on the topic, Wei-ho et al. [4], looks at identifying multiple singers in the same audio. The major challenges for this study arise from the fact that a singer's voice tends to be arbitrarily altered from time to time and is inextricably intertwined with the signal of the background accompaniment. Here, methods are presented to separate vocal form non-vocal regions and for isolating vocal characteristics from background music, and to distinguish singers from another. There is another research done to identify singers using the vocal characteristic; vibrato of a singer which is said to be unique in every singer [5].

In another effort in singer identification [6], the paper considers 32 Mel-Frequency Cepstral Coefficients in two subsets: the low order MFCCs characterizing the vocal tract resonances and the high order MFCCs related to the glottal wave shape. They explore possibilities to identify and discriminate singers using the two sets. Based on the results they can affirm that both subsets have their contribution in defining the identity of the voice, but the high order subset is more robust to changes in singing style. In another attempt [7] a methodology for Carnatic music singer identification is proposed and implemented. The motive behind identifying the singer is to extend this work for efficient music information retrieval of Carnatic music song based on singer identification. The features are based on Carnatic music characteristics like just tempered, varying pitch, varying interval of the octave, etc. The coefficients are called as Carnatic interval cepstral coefficients (CICC) since they are based on Carnatic music's octave interval. The input signal is segmented and frequency of each segment is determined which is followed by determining of the Carnatic interval cepstral coefficients. These coefficients are used to construct a GMM model which forms the basis for singer identification. There is also a novel scheme [8], called Hybrid Singer Identifier (HSI), for efficient automated singer recognition. HSI uses multiple low-level features extracted from both vocal and nonlocal music segments to enhance the identification process; it achieves this via a hybrid architecture that builds profiles of individual singer characteristics based on statistical mixture models.

A. *Acoustic Features of Singing*

There are many features that can be extracted from music signal. These features can be categorized into: reference features, content-based features and text-based features. A singing voice can be represented using content-based acoustic features which include timbral texture features, rhythmic content features and pitch content features. Based on these features, singing voice can be analysed and classified.

1. Pitch Features ([9] – [15])
2. Harmony Features
3. Formant Features [17]
4. Vibrato Features ([18] – [21])
5. Timbre Features ([22] – [25])

B. *Classifiers*

The purpose of classifier learning is to find a mapping from the feature space to the output labels by taking accurate decisions so as to minimize the prediction error. The common choices of classifiers are K-nearest neighbour, support vector machine, and GMM classifier.

1) *K-Nearest Neighbour (K-NN) Classifier*

The K-Nearest Neighbours algorithm is the simplest machine learning algorithm, which identifies the object by the majority vote of its neighbours based on distance (usually using a Euclidean distance). Given an input feature vector the algorithm finds k closest feature vectors representing different classes. The disadvantage of K-NN classifier is that its accuracy relies on the selection of an optimum number of neighbours and the most suitable distance measuring method. K-NN has been applied to various music sound analysis problems.

2) *Support Vector Machine (SVM)*

SVM is the state-of-the-art binary classifier based on the large margin principle and it works well with high- dimensional data. Intuitively, it aims at to construct a hyperplane that divides a data set into n regions, where n is the number of class labels in the data set. These hyperplane simplify to a set of Lagrange multipliers for each training case, and the set of points within the dimensional vectors fed for training that have non-zero Lagrangians are the support vectors. The machine saves these support vectors and applies them to new data in the form of the test set for further on-line classification. Therefore, the SVM has good classiûcation performance since it focuses on the difficult instances.

3) *Gaussian Mixture Model (GMM)*

The Gaussian mixture model uses multiple weighted Gaussians to attempt to capture the behaviour of each class of training data. The use of multiple Gaussians is particularly beneficial when analysing data that has a distribution not well modelled by a single cluster. It is known that GMMs provide good approximations of arbitrarily shaped densities of a spectrum over a long span of time, and hence can reflect the vocal tract configurations of individual singing voice.

DESIGN & METHODOLOGY

The overall system design is depicted in Figure 1. The system will mainly consist of a particular Dataset used for training and testing the system, a pre-processing unit, feature extraction unit, classification unit and other complimentary units which will help the whole system to carry out the Singer Identification task properly.

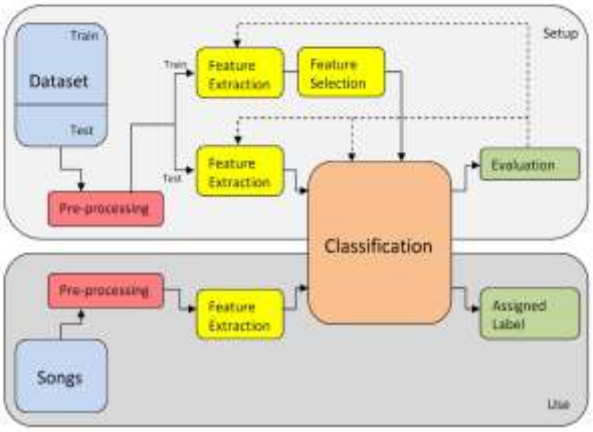

Fig. 1. Overall System Design

C. *Dataset*

Dataset is a set of Sri Lankan songs, representing a set of songs of well-established singers which they have sung within their period of active singing career. Hence, the songs were selected which represents the era of 1960s to 1990s.

D. *Pre-processing*

The pre-processing unit plays a vital part of the system as it will be transforming the songs and harnessing them to a format where feature extraction is possible. This process have several steps to it.

1. Converting the music files in different formats into WAV.It is always better to use a lossless format of data, therefore the music files were converted into WAV format.
2. Converting stereophonic audio files into monophonic by summing the two channels together.
3. Converting all the music audio files into one sampling rate of 44100Hz making the depth of all the music files into 32bits bit depth.
4. Songs were segmented in to two different sections via reliable frame selection; Vocal and Non-vocal segments, as shown in the Figure 2. Further processing will be done with the Vocal Segments for feature extraction.

A popular audio editor, Audacity[1] for this pre-processing task.

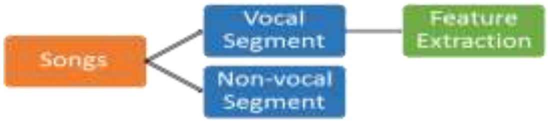

Fig. 2. Pre-processing

E. *Feature Extraction*

In this unit of the system, the Melody extraction from polyphonic music signals using pitch contour characteristics from Justin *et al.* [22] was used to extract the predominant frequency of the music audio files, which reduces the accompaniment sound (background music sound) to enhance the features of the voice. This was conducted using a Hann windowing function with a window size of 46.4ms, and a hop size of 2:9ms. The sequence of frequencies are then resynthesized to .wav format.

Then certain acoustic features such as Pitch, Harmony Features, Formant features, Vibrato Features and Timbre Features which are unique to the singers' voice will be identified and tagged for either training or for matching. These features are extracted as a 21 Mel Frequency Cepstral Coefficients (MFCCs), which are known to represent the distinct characteristics of the voice. Figure 3 illustrates the fashion in which the feature extraction process is conducted.

[1] http://audacityteam.org/

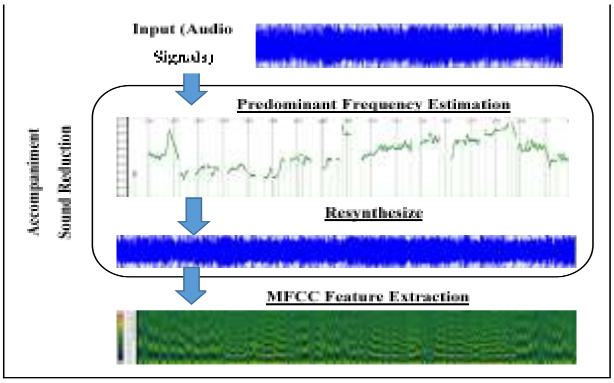
Input (Audio Signals)
Accompaniment Sound Reduction
Predominant Frequency Estimation
Resynthesize
MFCC Feature Extraction

Fig. 3. Feature Extraction Process

As shown in the above figure, the Accompaniment Sound Reduction is done to enhance the voice features. The Predominant Frequency calculation is done based on the voice. The following figure shows both the Input Audio Signal in blue and the Predominant Frequency in green.

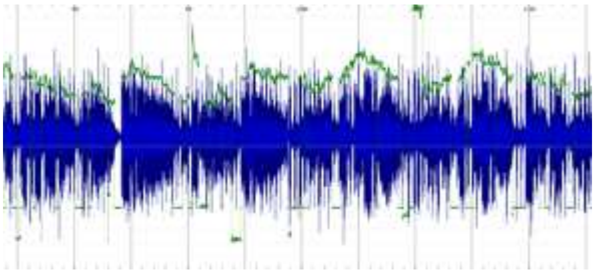

Fig. 4. Input Audio Signal in blue and the Predominant Frequency in green

MFCC Feature Extraction Process

Mel Frequency Cepstral Coefficients (MFCCs) are compact, short time descriptors of the spectral envelope audio feature set and typically computed for audio segments of 10-100ms. MFCC are one of the most popular set of features used in pattern recognition. MFCC was originally developed for automatic speech recognition systems, lately have been used with success in various musical information retrieval tasks. Although this feature set is based on human perception analysis but after calculated features it may not be understood as human perception of rhythm, pitch, etc. Normally first 13 MFCCs are used for musical information retrieval tasks. But this research uses 21 MFCCs for all calculations. Figure 4.5 illustrates the different steps in the calculation from raw audio signal to the final MFCC features.

1. The first step is dividing the speech signal into frames, usually by applying a windowing function at fixed intervals. The aim here is to model a small typically 20ms section of the signal, which are statistically stationary. The window function typically a hamming window and its removes edge effects.
2. Then for each short analysis window a spectrum is obtained using FFT.
3. In the next stage the Spectrum is passed through Mel-Filters to obtain Mel-Spectrum. This Mel band step is also a smoothing of the spectrum and a dimensionality reduction of the feature vector.
4. Cepstral analysis is performed on Mel-Spectrum to obtain MFCC. In the Cepstral analysis stage, take the logs of the value of the Mel bands and apply a set of Discrete Cosine Transform (DCT) filters on the Mel bands as if they were signals.

5. Finally the result is a lower dimensional feature of MFCCs. Thus music is represented as a sequence of Cepstral vectors and which are given to pattern classifiers for musical genre recognition purpose.

The voice is the major part of a song the human ear identifies, but when it comes as a single audio signal mixed with accompaniment sounds it makes it difficult to identify with a trained software. However, as the MFCC are proven to be somewhat unique to each and every individual which gives us the opportunity to identify the voice of the singer.

F. *Classification*

Classification will be conducted using the K-Nearest Neighbour Classifier, Support Vector Machine Classifier, Multilayer Perceptron Classifier, and Random Forest Classifier along with Principal Component Analysis feature section method. The feasibility and the performance of these methods are critically analysed in the next chapter.

EXPERIMENTAL RESULTS AND EVALUATION

This chapter will focus on elaborating the Experimental Results of the acquired dataset and the evaluation of the dataset comparing absolute and relative accuracy of each related experiment. These results will also include a critical analysis of each experiment and special cases noticed during the experiments.

The different combinations of the **Training datasets** are then used to train different models using different classification models using the open-source software, Weka. Different **classification algorithms** such as Support Vector Machine, Multilayer Perceptron, K-Nearest Neighbour and Random Forest were used in the classification, and Principal Component Analysis (PCA) is used as an optimization technique for feature selection.

All experiment reported in this research were performed using 10 fold cross **validation** method, in which the dataset was split into 10 parts of equal size. 90% of the songs were used to train the classifier and the rest 10% of the songs were used to test it. This was done 10 times, once for each fold, the classification is done for all the singer classes.

G. *List of Singers and Number of Songs used*

Table 1. List of Singers and Number of Songs Used

	Singer	**No. of Songs**
Male Singers	W D Amaradeva	21
	Victor Rathnayake	21
	Sunil Edirisinghe	21
	T M Jayarathne	22
	Mohideen Beg	11
	Ishaq Beg	16
	H R Jothipala	16
	Greshan Ananda	15
	Sunil Shantha	11
	Ivo Denis	10
	W D Ariyasinghe	21
Female Singers	Anjalin Gunathilake	21
	Nanda Malini	21

These numbers of songs were taken into consideration based on expert opinion and the number of songs were limited because of high memory requirement constraints. The bulk dataset classifications were subjected to bringing down the number of songs as well. However, with the given number of songs the experiments were very successful in classifying and identifying the singers up to a significant accuracy level.

When selecting the songs, they were selected representing;

- *all genres*
- *sung at different time periods of the singers' lifespan*
- *having different tempos*
- *having different accompaniment instruments.*

H. Significance of the vocal and non-vocal parts of songs on the performance of singer identification

According to Figure 5, it is clear that the non-vocal parts have a significant influence over singer identification in a larger scale. Hence it can be concluded that with only the vocal parts of the songs, identification of the singer can be done accurately. Hence, the subsequent experiments were carried out only with the vocal excerpts of the songs. Furthermore it can be advised that for a real world application on singer identification, that it's best that the only the vocal parts of the songs be used.

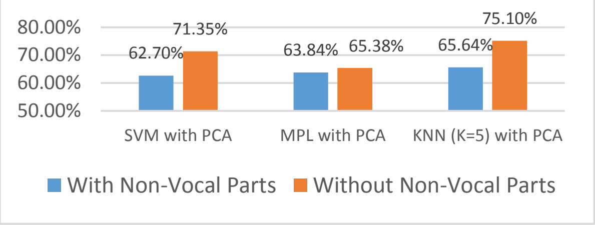

Fig. 5. Influence of Noo-Vocal parts of songs to the overall accuracy

I. Classification of Singers with Similar Voice

One of the main challenges the Singer Identification Research Problem faces in the field on Music Information retrieval is distinguishing between very similar voices. In this case, several experiments are conducted under this Experiment to critically evaluate on the performance of the MFCC features to accurately classify similar voiced singer. This experiments contains several couples of singers who are widely known to have similar voices within Sri Lankan music. Some of them are Father and Son; for example Mohideen Beg and Ishaq Beg. Hence, with all the test cases taken, it is believed this experiment is conducted up to the most extreme level of Singer Identification, and the results are very successful in terms of accuracy.

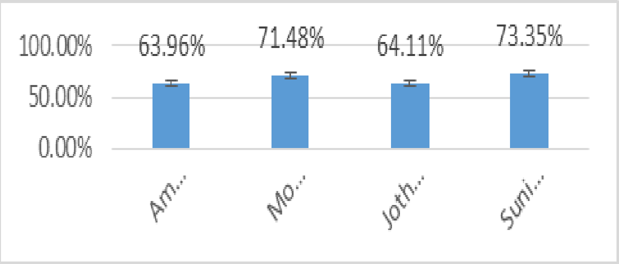

Fig. 6. Accuracies of Similar-Voiced Artist Classification

Figure 6 consists of the accuracies from the classification between four couples of similar-voiced artists. The best accuracies were achieved by the Random Forest classifier, and the accuracy obtained by KNN with Principal Component Analysis were only a few percentages lower than this. The test cases were as follows;

- Amaradeva vs. Ariyasinghe
- Mohideen Beg vs. Ishaq Beg
- Jothipala vs. Greshan Ananda
- Sunil Shantha vs. Ivo Denis

According to Figure 6, we can see that the proposed method gives out a good accuracy when the classification is conducted between similar-voiced artists.

J. *Identifying similar-voiced Singers when the classifier is only trained with one of the similar-voiced singer*

In this experiment, the objective is to observe whether a similar-voiced singer will be identified by the classifier, when the classifier is trained with the other similar voiced singer. The similar-voiced singers Mohideen Beg and Ishaq Beg are taken as the subjects for this experiments. The classifier model is trained with Mohideen Beg and three other singers; two male singers (Sunil and Victor) and one female singer (Nanda Malini) for better justification of the experiment. After the training, the model is re-evaluated with a test set consisting of Ishaq Beg and the classification results are observed. The experiment is done using three different classifiers; SVM, MLP and KNN with PCA to observe the accuracy across classifier over the experiment. The shows the confusion matrix of the classification conducted with the SVM classifier, which showed the most accuracy.

Table 2. SVM with PCA – Confusion Matrix

a	b	c	d	
0	0	0	0	| a = Nanda
0	0	0	0	| b = Sunil
0	0	0	0	| c = Victor
2689	3513	2565	**6490**	| d = IshaqBeg = MohideenBeg

K. *Classification with Increased Number of Singer Classes*

So far the singer identification classification was based on two-class classifications. This experiment is to observe the results of the classifications when the number of classes are increased. This was conducted as two experiments, first with 5 singer classes and the second experiment with 6 singer classes. Each sub-experiments were classified using SVM, MLP and KNN classifiers with PCA feature selection. The sub-experiments conducted were as follows;

1) *Singer Identification with 5 Singer Classes*

This experiment was conducted with 5 singer classes, namely two female singers; Nanda, Anjalin and three male singers; Sunil, Victor, Amaradeva. The overall accuracy figures of each classifier; SVM, MLP and KNN with PCA were as follows.

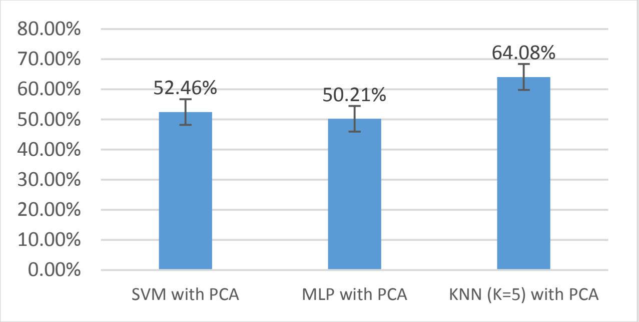

Fig. 7. Summary of Accuracies in 5-Singer class classification

2) Singer Identification with 6 Singer Classes

The above experiment is again conducted with one more male singer class (TM Jayarathne) added to the training data with the same three classifiers. The overall accuracy results were as follows;

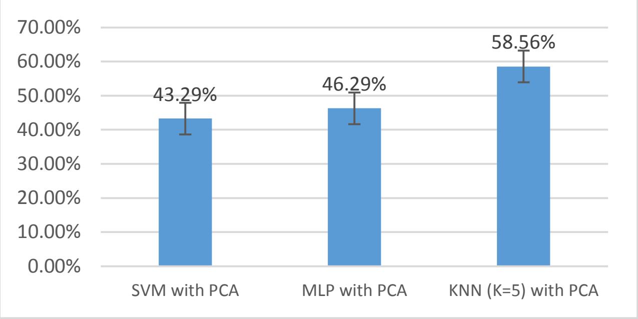

Fig. 8. Summary of Accuracies in 5-Singer class classification

According to both sub-experiments conducted under *E*, we can see that with the increasing number of classes the overall accuracy is gradually decreasing as depicted in the Figure 9. However the classification of singer classes is still capable without any doubt. Hence the proposed research solution on Singer Identification is possible.

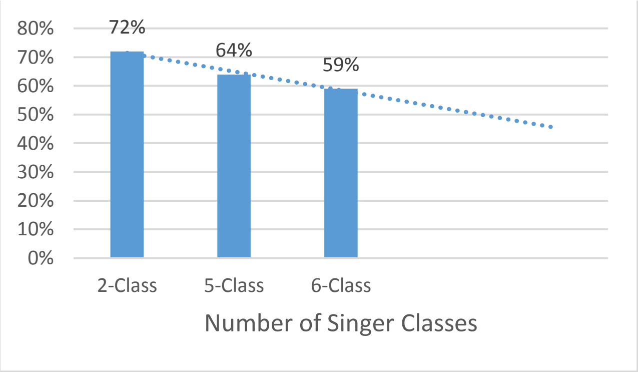

Fig. 9. Projection of Accuracies against the Number of Classes

L. *Classification using the first 13 MFCCs*

According to the literature written on speaker identification research, it is said, that out of the MFCC's extracted from voice signals, the first 13 MFCCs are the most significant in terms of features in disfurnishing ones voice with higher accuracy and performance. Hence, this experiment is conducted to observe whether Singer Identification can be optimized in terms of accuracy and performance within the proposed research solution when only the first 13 MFCCs are used as opposed to the 21 MFCCs used in this research as the feature vector.

This experiment consists of two sub-experiments which focuses on two different scenarios, on which experiments were conducted earlier in this research. The two sub-experiments are;

1) *Amaradeva Vs Ariyasinghe classified using only the First 13 MFCCs*

This sub-experiment is conducted to observe how accuracy and performance would be improved in a 2 – Class Singer Identification classification model when only the first 13 MFCCs are used as the features as opposed to the 21 MFCCs used in this research.

As displayed in Figure 10, it is clear that using only the first 13 MFCCs as the feature vector has made a slight adverse effect on the classifier performance and the accuracy, but the magnitude of the adverse effect varies with the classifier used for classification.

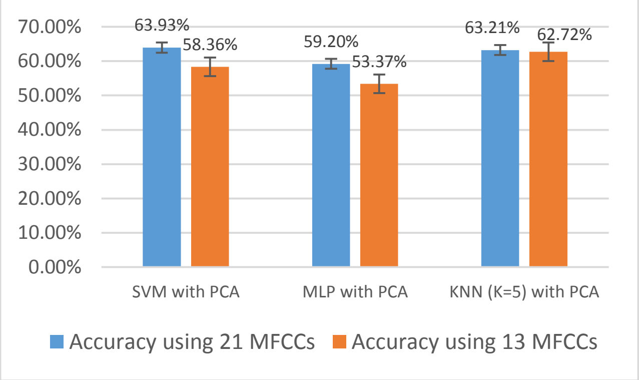

Fig. 10. Compared Accuracies of Classification Results using 21 and 13 MFCCs

2) *Class Multiple class classification using only the first 13 MFCCs*

This sub-experiment is conducted to observe how accuracy and performance would be improved in a 6 – Class Singer Identification classification model when only the first 13 MFCCs are used as the features as opposed to the 21 MFCCs used in this research.

As displayed in Figure 11, it is clear that using only the first 13 MFCCs as the feature vector has made a slight adverse effect on the classifier performance and the accuracy, but the magnitude of the adverse effect varies with the classifier used for classification.

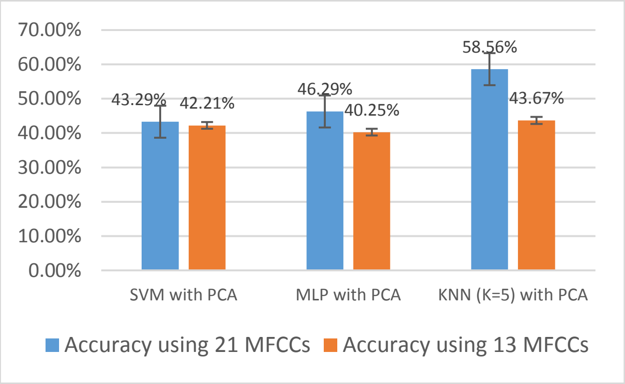

Fig. 11. Compared Accuracies of Classification Results using 21 and 13 MFCCs

It can be summarized that using the first 13 MFCCs as the feature vector, **is not a viable option** when classifying for Singer Identification.

M. Singer Identification in Duet Songs

This experiment was conducted to observe if the proposed research solution is able to identify singers in duet songs. Duets sung by Jothipala and Anjalin was taken as the test set and they were re-evaluated with a KNN classifier trained with four singers (with two male singers; Jothipala and Greshan, and two female singers; Anjalin and Nanda). KNN with PCA model was used as it was the highest performing classifier for multiple class classifications.

Table 3. Duet Singer Identification - Confusion Matrix

1784	|	a = Jothi
881	|	b = Greshan
457	|	c = Nanda
2131	|	d = Anjalin

As displayed in Table 3, it is clear that the Jothipala and Anjalin were correctly classified up to a satisfactory level to be identified as the singers in the duet songs. Hence, it can be concluded that this research solution is able to identify singers of a duet songs. This could be also taken up to more than 2 singers singing in the same song, but depending on the amount of data generated when extracting the features, the results may differ.

DISCUSSION & CONCLUSION

The objectives of this research was to provide research solutions to find ways for singer identification in polyphonic music, find ways to correctly identify and distinguish between very similar singing voices, improve or find ways for feature extraction & feature selection, and to find the best supervised learning approaches in different scenario of singer identification. Looking at the experimental results and the evaluation, it is evident that this research has contributed to finding solutions to all of the research objectives up to a prominent level, and generate more research evidence and knowledge on Music Information Retrieval.

Main difficulty in modelling the characteristics of a singing voice in polyphonic music lies in the negative influence of accompaniment sounds. Addressing that problem, this research focused on accompaniment sound reduction through predominant frequency generation and resynthesizing of the music files, which contributed to the accuracy and the performance of singer identification in very large scale. Subsequently, selection of vocal-only parts of the songs for the classification enabled in increasing accuracy and performance in overall singer identification and distinguishing between very similar singing voices.

In terms of features and feature extraction, the feature vector used in this research; 21 Mel Frequency Cepstral Coefficients (MFCCs), which represented complex voice characteristics in a very rich and subtle manner played a vital role in the successful singer identification in the research. MFCC extraction, which was a five step process performed on a windowed excerpt of the music signal extracted the voice features of the predominant frequency of the original music. It can be said that slight variances of the window size might cause the richness of the features extracted, but this research used a default window size constantly over the research, which was believed to be the optimal window size to extract voice base features.

The different Experiments, carried out covered most of the expected scenarios expected out a Singer Identification System and the results were very good in terms of accuracy and performance. One of the difficult tasks of a singer identification system is distinguishing and identifying similar singing voices, which was successfully resolved by the approach proposed by this research. Furthermore, this approach is capable of singer identification with multiple singer classes as well, where it would be most relevant to a real world singer identification or an audio monitoring system. Subsequently, this research has also addressed a real time application problem of identifying singers of songs sung by more than one individual, such as duets, where the results were accurate up to an acceptable level.

Concluding the research discussion, it can be said that the proposed approach in addressing the problem of Singer Identification and its' related sub-problems were successfully addressed. However, there can be endless possibilities to improve accuracy and performance of this approach and find ways to fill in the loopholes of this approach. Hence there is lot of room for future work.

FUTURE WORK

As discussed in the conclusion, there are many possibilities for improvements to be made in the approaches taken to provide a solution for the problem of Singer Identification and its' related sub-problems. With the work carried out by this research, much knowledge on Music Information Retrieval was gathered and some of the most potential future work were identified along the path.

In terms of the Features, there are other derivations of the MFCC, such as LPC (Linear Predictive Coding)-Derived Mel Cepstral Coefficients (LPMCCs) that has been used for speaker identification which can be used for singer identification, where increased accuracy may be expected. Furthermore, as discussed in the conclusion it is also possible that slight deviations of the window size taken in extracting the MFCCs may cause different outcomes.

In terms of classification approaches, there could be other classification algorithms that gives out higher performance and accuracy, and more feature selection approaches such as classification subset evaluation can be used to select the most optimal sub set of the features in the classification.

Looking at the real world application side of Singer Identification, approaches can be tried out in automating this research approach for ease of use, and different variations of applications can be implemented for the betterment of the overall music industry.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my supervisor Dr Lakshman. Jayaratne for the continuous support and the supervision provided throughout the study. Furthermore, I would like to thank Mr. Suresh Maliyadda, an outstanding music director, composer and one of the most experienced musicians in the current Sri Lankan music field for his expert opinion, and last but not least, my colleagues of the Music Information Retrieval research group.

REFERENCES

[1] H. Fujihara, M. Goto, T. Kitahara and H. Okuno, "A Modeling of Singing Voice Robust to Accompaniment Sounds and Its Application to Singer Identification and Vocal-Timbre-Similarity-Based Music Information Retrieval," in *Audio, Speech, and Language Processing, IEEE Transactions*, 2010.

[2] A. K. T. V. Toni Heittola, "Musical instrument recognition in polyphonic audio using source-filter model for sound separation," 2009.

[3] T. Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria," *IEEE Transactions on Audio, Speech & Language Processing - TASLP no. 3, pp. 1066-1074,*, vol. 15, no. 3, pp. 1066-1074, 2007.

[4] H.-m. W. Wei-ho Tsai, "Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals," *IEEE Transactions on Audio, Speech & Language Processing - TASLP*, vol. 14, no. 1, pp. 330-341, 2006.

[5] H. L. Tin Lay Nwe, "Exploring Vibrato-Motivated Acoustic Features for Singer Identification," *IEEE Transactions on Audio, Speech & Language Processing* , vol. 15, no. 2, pp. 519-530, 2007.

[6] J. A. Annamaria Mesaros, "The Mel-Frequency Cepstral Coefficients in the Context of Singer Identification," in *International Symposium/Conference on Music Information Retrieval - ISMIR , pp. 610-613,* 2005.

[7] T. V. G. Rajeswari Sridhar, "Music Information Retrieval of Carnatic Songs Based on Carnatic Music Singer Identification," in *International Conference on Computer and Electrical Engineering - ICCEE , ,* 2008.

[8] J. S. B. C. K.-I. T. Jialie Shen, "A novel framework for efficient automated singer identification in large music databases," *ACM Transactions on Information Systems - TOIS,* vol. 27, no. 3, pp. 1-31, 2009.

[9] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” IEEE Trans. Speech Audio Process., vol. 10, no. 5, pp. 293–302, 2002.

[10] T. Li and M. Ogihara, “Detecting emotion in music,” in Proc. Int. Conf. Music Information Retrieval, 2003.

[11] T. Fujishima, “Realtime chord recognition of musical sound: A system using common lisp music,” in Proc. Int. Computer Music Conf., 1999, pp. 464–467.

[12] E. Gomez, “Tonal description of music audio signals,” Ph.D. dissertation, Dept. Technol., Universitat Pompeu Fabra, Barcelona, Spain, 2006.

[13] J. Serra, E. Gomez, P. Herrera, andX. Serra, “Chroma binary similarity and local alignment applied to cover song identification,” IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 6, pp. 1138–1151, 2008.

[14] M. Marolt, “A mid-level melody-based representation for calculating audio similarity,”in Proc. Int. Conf. Music Information Retrieval, 2006.

[15] G. Poliner, D. Ellis, A. Ehmann, E. Gomez, S. Streich, and B. S. Ong, “Melody transcription from music audio: Approaches and evaluation,” IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 4, pp. 1247–1256, 2007.

[16] P. R. Cook, “Identification of control parameters in an articulatory vocal tract model, with applications to the synthesis of singing,” Ph.D., Stanford Univ., CA, 1990.

[17] Sundberg, J., “The acoustics of the singing voice,” Scientific American, vol. 236, pp. 82–91, 1977.

[18] D. Gerhard, “Pitch-based acoustic feature analysis for the discrimination of speech and monophonic singing,” J. Canadian Acoust. Assoc., vol.30, no. 3, pp. 152–153, 2002.

[19] Nwe, T.L., Li, H.: Exploring Vibrato-Motivated Acoustic Features for Singer Identification. IEEE Transactions, Audio, Speech and Language Processing 15(2) (2007)

[20] Wei-Ho Tsai, Member, Hsin-Chieh Lee “Automatic Evaluation of Karaoke Singing Based on Pitch, Volume, and Rhythm Features” IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 4, pp. 1233- 1243, May 2012

[21] Sundberg, J., “Human singing voice,” in Encyclopedia of Acoustics,pp.1687–1695, John Wiley and Sons, Inc., 1997.

[22] Justin Salamon and Emilia Gómez. “Melody extraction from polyphonic music signals using pitch contour characteristics”. In: Audio, Speech, and Language Processing, IEEE Transactions on 20.6 (2012), pp. 1759–1770.

[23] L.Macy, Grove Music Online. [Online]. Available: http:www.oxford-musiconline.com/public/book/omo_gmo.

[24] Cleveland, T.F.: Acoustic Properties of Voice Timbre Types and Their Influence on Voice Classification. Journal of Acoustical Society of America 61, 1622–1629 (1977)

[25] Poli, G.D., Prandoni, P: Sonological Models for Timber Characterization. Journal of New Music Research,170–197 [25] Zhang, T., Kuo, C.C.J.: Content-Based Audio Classification and Retrieval for Data Parsing. Kluwer Academic Publishers, USA (2001)