Guitar Model Recognition from Single Instrument Audio Recordings

David Johnson
Department of Computer Science
University of Victoria
Victoria, BC
Email: davidjo@uvic.ca

George Tzanetakis
Department of Computer Science
University of Victoria
Victoria, BC
Email: gtzan@uvic.ca

***Abstract*—The main goal of this paper is to explore the recognition of particular guitar models from single instrument audio recordings. This is different than existing work in music instrument recognition that deals with identifying different instrument types. Through a set of experiments we evaluate different sets of audio features and classifiers for this purpose. To improve accuracy a composite classifier is implemented to first discriminate between electric and acoustic guitars. This affords flexibility in training different models for each guitar type. A data set consisting of audio recordings from 15 guitar models, each recorded with a set of different playing configurations, is used for training and testing. We have found that K Nearest Neighbors and Support Vector Machine (SVM) classifiers perform the best. Testing is done by leaving a specific playing configuration out of the training model. Specific test cases show satisfactory results, with one test case achieving over 70% accuracy and a second one over 50%; both using a composite SVM model.**

I. Introduction

Music instrument recognition is process of identifying specific instruments found in an audio signal. Early work in music instrument identification was focused on detecting instruments in audio signals with one instrument. While more recently, research has moved to detecting instruments in recordings with multiple instruments. Our research explores a subset of instrument recognition of single instrument audio recordings, namely the identification of specific guitar models in a single instrument audio recording. Using a set of experiments we explore the optimal features and classification models for identifying a specific guitar model from a set of guitar recordings.

Music instrument identification has been a widely explored topic in the past couple of decades. One of the first attempts at instrument classification is [1]. In this research, the authors build a classification system to distinguish between 8 different instruments. They experimented with three different features, Linear Prediction Coefficients (LPCC), Cepstral Coefficients and Mel-Frequency Cepstral Coefficients (MFCC) and two classification algorithms, Gaussian Mixture Models and Support Vector Machines (SVM). The authors found the best results using the SVM, with a one-vs-all approach for handling multiple labels, and the MFCC feature set.

Agostini et. al. evaluate a set of spectral features and classification algorithms for music instrument recognition from a monophonic audio signal [2]. Through this work the authors found that Support Vector Machines performed poorly on stringed instruments. However their feature selection ignores the time-varying nature of tones.

Essid et. al. evaluate a set of pairwise classification techniques instead of the more common one vs all approach used with binary classifiers [3]. They also use a unique approach to feature reduction using Genetic Algorithms and Inertia Ratio Maximization using Feature Space Projection (IRMFSP). This approach removes the need to extract all features before classification.

More recently researchers have been working on the harder problem of detecting instruments from mixed audio signals with multiple instruments. Barbedo and Tzanetakis propose a technique for instrument recognition in mixed audio signals using partials [4]. Eggink and Brown present a system of identification of an instrument in the presence of an accompaniment [5].

Although music instrument detection is still an open problem, to our knowledge, little research has been done to detect specific models or characteristics of a particular instrument. There are, however, a few examples of highly specific classification tasks thare are closely related. Brown et. al. examine features and classification strategies for identification of woodwind instruments [6] and Fragoulis et. al. explore a novel approach for discriminating between piano and guitar using nontonal information. [7].

An understanding of timbre is important for instrument identification, especially in the cases where timbral differences are minimal, such as identification of instruments within a instrument family. Timbre is the characteristic of sound that allows listeners to distinguish it from another sounds independent of the the pitch and intensity. While other audio characteristics such as pitch and loudness are unidimensional, timbre is composed of temporal and spectral features. The guitar family has a set of timbral characteristics that set it apart from other instrument families such as woodwinds, for example. On the other hand, detecting a specific guitar from a set of guitars may rely on subtle differences in timbral features.

Many early studies of timbre, use multidimensional scaling techniques based on experiments with human participants to better understand the perceptual similarities between tones of different instruments [8]. Using these techniques Gray found three dimensions of timbre; spectral energy distribution, the energy in attack and, the synchronicity of transients in higher harmonics. Iverson and Krumhansl perform perceptual

978-1-4799-7492-4/15/$31.00

TABLE I. GUITAR MAKES AND MODELS

Make and Model	Type Label	Label
Norman B-T5-12	Acoustic	a1
Washburn WP-26-SENS	Acoustic	a2
Rainsong BT-WS1000N2	Acoustic	a3
Composite Acoustics X HG	Acoustic	a4
Composite Acoustics Cargo Raw	Acoustic	a5
Riversong Trad-CDN-SE	Acoustic	a6
Gibson L-4 Archtop	Acoustic	a7
Veelah V5.6 OMCE	Acoustic	a8
Beaver Creek BCTD 901 CE	Acoustic	a9
Paco Castillo 204	Acoustic	a10
Veelah ToGo-S	Acoustic	a11
Veelah ToGo-S	Acoustic	a12
Veelah ToGo-S	Acoustic	a13
Composite Acoustics Cargo Raw	Acoustic	a14
New York Pro 335	Electric	e1
Gibson Les Paul SE	Electric	e2
Fender Stratocaster	Electric	e3

experiments to find which dynamic attributes contribute to the timbre of a musical instrument [9]. Their experiments found that perceived similarity was due to centroid frequencies and amplitude envelopes of the tones, indicating that a combination of spectral and temporal characteristics are important for timbral characterization.

Some recent studies of timbre have attempted to understand it through computational methods, as opposed to using human subjects. Cosi et. al. cluster timbres using MFCC and a self-organizing neural network [10]. They argue that the steady state information from timbre is sufficient for analyzing timbre quality. Poli and Prandoni use computational methods to model the timbre space similar to the space found using MDS techniques from Gray [11]. Using an MFCC feature set the authors perform data analysis using PCA and self-organizing neural networks to derive a timbral space. They also argue that steady state information is sufficient for analyzing timbral quality but temporal features are needed for recognition.

II. DATA COLLECTION

For the classification experiments we've generated a dataset consisting of different guitars models with varying characteristics such as shape, size and material. The guitar samples were recorded over two sessions with each recording session occurring in a different room. In the first recording session, we recorded samples from a selection of 15 guitars, 12 acoustic and 3 electric, and recorded four different configurations from each to generate a dataset containing a variety of timbral characteristics describing each guitar's timbral signature. The configurations recorded include each open string played individually (A), a chord progression (B), a finger picking pattern (C), and a melody (D).

In the second recording session we used the same 12 acoustic guitars (although we had to replace one of the Veelah ToGo-S models with a new one, as the previously recorded guitar had been sold) plus the addition of a second Composite Acoustic Cargo Raw. During this session we recorded two different configurations. One configuration being the same chord progression from session one (E). The second configuration is another chord progression with chords played from the top of the neck to the bottom meant to represent the full chromatic range of each guitar (F).

In total we recorded 17 guitars which were selected to have different characteristics that attribute to different timbral signatures. See Table I for a list of guitars models recorded for the dataset. Two sets of the acoustic guitars selected were of the same guitar model, A11, A12 and A13 are all of the same model and A5 and A14 are of the same model. It would have been ideal to have multiple pairs of guitars with the same model but we were constrained to the guitars available at the guitar store. After the recording sessions, each recording was reviewed and processed to remove silence and speech at the beginning and ends of the recordings. Recordings were then annotated with two sets of ground truth labels. The first ground truth label is the the guitar type, acoustic or electric, and the second is a label indicating which guitar it is a recording of, see Table I for a list of guitar types and labels.

III. FEATURE EXTRACTION

Timbre, as previously discussed, is a combination of spectral and temporal features of an audio signal. To better understand the timbral characteristics of a guitar we explore different sets of features for use in guitar model classification. A set of potential features for music instrument recognition is presented in [12] and [13]. The field of speech recognition also provides a potential features for use, as timbre is an important quality of speech. Features that have proven useful in speech recognition are Linear Prediction Coefficients (LPCC) [14] and Mel-Frequency Cepstrum Coefficients (MFCC) [15]. While generally used for speech recognition, linear prediction has been used in instrument detection in [2] and [12]with varying success. MFCC were also used in [2] and [12] with better success. Spectral features may also prove useful as MFCC has shown to be insufficient for some instruments in [16]. An easy to extract, yet robust set of features instrument detection is presented [3].

With these features in mind, we create four different features sets to use for evaluating a set of guitar model classifiers. These include

1) LPCC,
2) Timbral features including MFCC, Zero Crossing Rate (ZCR), Centroid, and Flux,
3) MFCC Only and
4) Timbral features with Chroma.

For each of these feature sets we perform extraction using the bextract utility from Marsyas [17]. Features are extracted from each recording using a window size of 512 samples which are then aggregated to obtain statistical texture window. Each feature vector is composed of the means and variances of each feature, calculated over a sliding window of 20 extraction windows.

IV. CLASSIFICATION

For guitar model classification we experiment with three standard classifiers, Support Vector Machines (SVM) [18], Bayesian Classifiers [19] and k-Nearest Neighbors (kNN) [20].

TABLE II. EVALUATION RESULTS USING INDIVIDUAL FEATURE VECTORS FOR RECORDING SESSION 1

		Timbral with Chroma	Timbral	MFCC Only	LPCC
Baseline	10FCV	8.41%	8.41%	8.41%	8.41%
Naive Bayes	10FCV	24.60%	34.19%	35.00%	14.82%
	A12 Test	12.57%	17.90%	***19.91%***	13.24%
SVM	10FCV	59.66%	55.10%	51.69%	22.31%
	A12 Test	***30.36%***	***29.31%***	***30.11%***	***18.51%***
KNN	10FCV	99.93%	99.94%	99.93%	68.64%
	A12 Test	***37.54%***	***37.93%***	***36.77%***	***14.43%***

These classifiers were selected because they have been shown to perform well for instrument recognition tasks in [2], [3], [1], [21] and [22]. An SVM was shown to outperform a Gaussian Mixture Model in [2]; however, stringed instruments, such as guitars, were the most misclassified. In [3] an SVM was implemented using a pairwise classification strategy, instead of a one-vs-all approach to improve classification accuracy. Bayesian classifiers were used in both [21] and [22] with success. K-NN was also shown to have success with [22]. Another benefit of k-NN is its lazy learning capabilities afford classification models that can quickly learn from new input without having to retrain the model before querying for prediction.

V. EXPERIMENTS

A. Design

To find an optimal combination of features and classification methods we perform a set of experiments to test each feature set classifier pair. Using the sets of features and classifiers previously discussed, we evaluate each possible feature set and classifier pair using 10-fold cross validation (10FCV). We've also included a classifier that makes predictions based on the most frequent class to gather baseline statistics. To limit the number of variables contributing to classification we perform separate classification experiments for each recording session. In addition to cross validation we also evaluate the classification models using a training set that contains each guitar model but excludes the additional guitars of the same model. For example, in the first recording session guitar A12 is the same model as A11. By excluding guitar A12 from the training set and labeling it as A11 in our test set we can test if the classifiers are able to identify a guitar model (not just the specific guitar). We perform the same set of experiments for the guitar recordings from the second recording session. This time we exclude A13, the same model as A11, and A14, the same model as A5, from the training data to use as individual test data sets, similar to the A12 test.

B. Results

Tables II and III provide the evaluation results for each feature set and classifier pair using cross validation and the individual guitar model tests. The results are based on prediction of each feature vector in the test set. Highlighted results of the individual guitar model tests indicate the the majority of feature vector predictions are of the correct label, thus indicating a majority voting technique would classify this guitar correctly.

TABLE III. EVALUATION RESULTS USING INDIVIDUAL FEATURE VECTORS FOR RECORDING SESSION 2

		Timbral with Chroma	Timbral	MFCC Only	LPCC
Baseline	10FCV	9.53%	9.53%	9.53%	9.53%
Naive Bayes	10FCV	34.64%	46.94%	53.34%	14.62%
	A13 Test	***20.10%***	***54.97%***	***24.70%***	***49.30%***
	A14 Test	12.13%	**20.97%**	4.84%	14.08%
SVM	10FCV	81.08%	73.62%	70.08%	23.15%
	A13 Test	***65.83%***	***47.90%***	***44.60%***	***30.08%***
	A14 Test	14.42%	15.19%	11.76%	18.31%
KNN	10FCV	99.96%	99.97%	99.96%	70.54%
	A13 Test	***55.04%***	***42.74%***	***41.59%***	***22.41%***
	A14 Test	15.73%	13.61%	5.75%	***16.03%***

While all results are an improvement over the baseline, the SVM and kNN classifiers perform the best for all feature sets. With near 100% accuracy, the results of cross validation for kNN appear to suffer from high overfit. Overfit is most likely due to the nature of our feature extraction process combined with cross validation. Since the feature vectors represent statistical windows, neighboring vectors will have very similar statistics. Using cross validation, the distribution of neighboring vectors into both the training and test sets is a probable cause for overfitting in all experiments. K-NN, however, still shows promising results with the high accuracy values for the A12 and A13 test sets. The classification of guitar A14 has less than desired results. While A14 is the same model as A5, it was a used guitar, around two years old, and had a different set of strings. These factors were most likely the leading cause in misclassification. Also notice that with the top two classifiers there is little variation in the accuracy for the different feature sets with the exception of LPCC which are the lowest performing features overall.

The results of these experiments indicate that guitar models do have unique timbral qualities allowing standard classification models to differentiate between different models. In this set of experiments, however, we are simply evaluating the classification of individual feature vectors leading to overfit as previously discussed. A more realistic classification model would make predictions based on segments of audio. The results of these experiments are used to guide our implementation of a guitar model classifier that classifies the guitar recordings based on segments of audio of a defined length, discussed in section VI.

VI. AUDIO SEGMENT CLASSIFICATION

A. Classification Model Design

The results of our initial experiments are promising and show that there is potential for guitar model identification. These results, however, are based on individual feature vector classification. To accurately predict a segment of audio we need an implementation that aggregates results over a period of time. To accomplish this we have implemented a guitar model classifier that takes a WAV file or a music collection file as input, predicts the guitar model for one second segments of the recordings and outputs the results. Figure 1 presents an architectural overview of this guitar model classifier.

In our initial experiments we generated models using train-

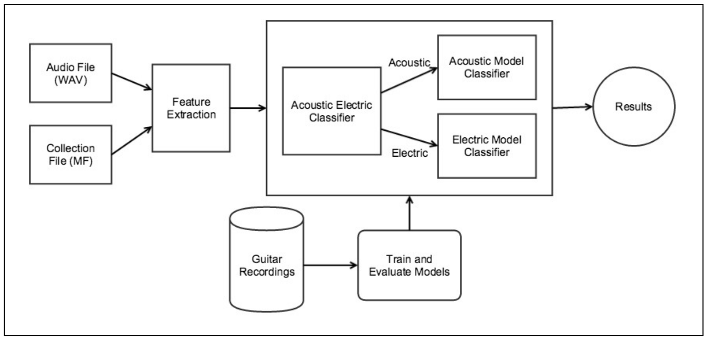

Fig. 1. Guitar Model Classifier

ing sets that contain both acoustic and electric guitars. While confusion matrices show only some overlap between these types of guitars, we feel that classification of guitar models will benefit from classification models that treat each guitar type distinctly. As such, we have implemented a hierarchical classifier that first predicts a guitar type of acoustic or electric. Based on this prediction then predicts guitar model using a classification model tuned specifically for the predicted guitar type.

Each model is generated by treating the training data as a bag of features. In other words, all individual feature vectors from a recording are labeled with the ground truth and then added to the model. The model is then trained using a specified classifier. Classifying individual feature vectors, however, is not sufficient when trying to predict the guitar model from an audio recording. Instead prediction should be based a segment of the recording. To accomplish this we implement prediction using majority voting based on a defined length of audio (the default segment length is one second). In other words, we predict the label for each feature and accumulate the results for a defined segment length. The label with the most votes wins as the prediction for that time segment.

B. Evaluation

Defining a strategy for partitioning the dataset into training and tests is important to validating the generalizability of the guitar model classifier. To avoid overfit we use a strategy that partitions the data based on recording configuration. To create our training sets we exclude all feature vectors of one configuration and use those for the test set; one training set, for example, contains the feature vectors from configurations A, B, and D and the corresponding test set contains feature vectors of configuration C. To better understand how different recording rooms affect recognition we first perform these tests for each recording session. Additionally, we generate training sets containing all the configurations for each session but excludes the individual guitars of which we have multiple models, namely A12, A13, and A14, similar to our previous experiments. Finally, we test the guitar model classifier with recordings from both sessions by generating a training set of all configurations from recording session one and a test set using feature vectors from configuration E. Configuration E is the same chord progression as configuration B from the first session.

Using these strategies we ensure that our test sets include feature vectors that are independent from the training set.

TABLE IV. EVALUATION RESULTS USING AUDIO SEGMENTS OF RECORDING SESSION 1

		Timbral	**MFCC Only**
Baseline	A:	6.48%	4.86%
	B:	7.02%	6.58%
	C:	6.74%	5.06%
	D:	5.63%	8.13%
	A12:	7.79%	7.79%
SVM	A:	20.24%	19.84%
	B:	57.45%	54.82%
	C:	73.88%	73.60%
	D:	25.63%	26.88%
	A12:	53.25%	50.65%
kNN	A:	17.81%	17.81%
	B:	47.80%	47.80%
	C:	59.55%	59.55%
	D:	20.63%	20.63%
	A12:	50.65%	50.65%

TABLE V. EVALUATION RESULTS USING AUDIO SEGMENTS OF RECORDING SESSION 2

		Timbral	**MFCC Only**
SVM	E:	62.05%	62.65%
	F:	61.82%	60.91%
	A13:	79.41%	79.41%
	A14:	16.67%	16.67%
kNN	E:	53.01%	53.01%
	F:	55.00%	55.00%
	A13:	79.41%	79.41%
	A14:	8.33%	8.33%

In this case, the test set feature vectors are from separate recordings with varying of timbral qualities. Because each recording configuration has different timbral qualities, this strategy also affords us an analysis of how the different qualities may contribute to the overall timbral signature of a guitar. Partitioning our training and test sets in this manner will help provide insight into the possible generalizability of the models.

For evaluation, we decided to train the guitar model classifier's component models using both kNN and SVM classifiers since these performed the best in our initial experiments. For features we test with the timbral feature set and the MFCC only feature set. Although the timbral with chroma feature set performed well we didn't include it in our testing because the additional chroma features don't appear to add value to our classification results. Our initial evaluation of the guitar model classifier trains each component model using the same feature set and classifier. This leads to four different testing configurations for each feature set and classifier pair. For baseline statistics we've also included a dummy classifier using a stratified prediction strategy.

TABLE VI. RESULTS OF TESTING THE EFFECT OF RECORDINGS FROM DIFFERENT ROOMS

	Timbral	MFCC Only
Baseline	7.23%	7.23%
SVM	24.10%	24.10%
kNN	23.49%	23.49%

C. Results

Evaluation results for each configuration can be found in Tables IV, V and VI. Table IV presents the results of the first recording session. In this case, the performances of each configuration are quite similar through all tests. Interestingly, the order of performance for each recording configuration is consistent through all classification configurations, such that C and B are always the top performers. This gives us some indication that the selected feature sets contain important information about a guitar's timbral signature. Overall, the SVM performs better for all recording configurations with a significant increase in accuracy for configuration C. In both classification configurations the added features in the Timbral feature set don't add much value to classification. Using the MFCC only feature set appears to be sufficient for classification in these cases. The caveat is with the A12 test set where the timbral feature set slightly improves performance when using an SVM for classification.

Table V shows classification results for the second recording session. In this case we have two configurations so we trained our models with one configuration and tested with the other. In both test cases, we have similar results across the classifiers. As with tests from recording session one, both classifiers perform adequately with the SVM performing slightly better. Generating a test set from guitar A13 (the same guitar model as A12) also provides similar results to experiments from recording session one. Using a test set generated from A14, on the other hand, shows that other variables contribute to the recognition of guitar models. In this case A14 is a used version of the same model as guitar A5, approximately two years old, and strung with a different set of strings.

Table VI shows the results from performing classification on a test set of guitar recordings from a different room. In this case we generated the training set using all configurations from recording session one and the test set was generated using configuration E from the second recording session. We chose this configuration since it is the same playing pattern of configuration B which is included in our training data. Our results show improvement over the baseline but do not have the same level of accuracy when testing recordings of the same room. This provides some indication that our features capture some timbral qualities of the guitars but that recording room has an effect on the recognition.

Table VII presents a confusion matrix for the guitar model classifier using a prediction size of one second of audio and trained with SVMs and timbral features. This matrix was generated using recording configuration B for testing. Reviewing this confusion matrix we notice two insights. First many of the audio segments for each guitar model are often misclassified as guitar A1. The second insight is that a high percentage of electric guitar segments are being classified as

TABLE VII. CONFUSION MATRIX FOR THE GUITAR MODEL CLASSIFIER CONFIGURED WITH SVMS AND THE TIMBRAL FEATURE SET USING A RECORDING CONFIGURATION B FOR TESTING

a1	a10	a11	a2	a3	a4	a5	a6	a7	a8	a9	e1	e2	e3	
16	0	0	0	0	0	0	1	0	0	0	0	0	0	**a1**
0	15	0	0	0	0	0	0	0	0	0	0	0	0	**a10**
0	0	14	0	1	0	0	0	0	1	0	0	0	0	**a11**
4	0	6	5	0	0	0	0	0	0	0	0	0	0	**a2**
4	0	0	0	4	0	2	1	1	5	0	0	0	0	**a3**
6	0	0	0	0	10	0	0	0	0	0	0	0	1	**a4**
3	0	0	0	0	1	11	0	0	0	2	0	0	0	**a5**
1	0	0	0	0	0	0	14	0	0	1	0	0	0	**a6**
3	0	1	0	0	0	0	2	8	2	0	0	0	0	**a7**
2	0	2	0	0	0	0	0	0	11	2	0	0	0	**a8**
1	6	0	0	0	0	0	0	0	4	5	0	0	0	**a9**
0	0	0	0	0	0	0	6	0	0	0	7	5	0	**e1**
0	2	0	0	0	0	0	7	0	0	0	0	6	1	**e2**
1	0	0	0	0	2	0	6	0	1	0	0	0	5	**e3**

TABLE VIII. PREDICTION RESULTS FOR THE A12 TEST USING THE CLASSIFIER TRAINED WITH SVMS AND TIMBRAL FEATURES

a1	a10	a11	a2	a3	a4	a5	a6	a7	a8	a9	e1	e2	e3	
2	1	41	6	2	1	1	0	1	20	0	1	0	1	**a12 ToGo-S**

acoustic. To prevent this we will need to better tune our acoustic/electric classifier.

The results A12 test for the same classification model just discussed, in Table VIII. In this case the high majority of the predictions are spread between two guitars. The majority of the segments are predicted correctly as A11, the other Veelah ToGo-S. The second most predicted guitar is A8 another Veelah guitar. This indicates that these two guitars may have similar timbral qualities.

VII. CONCLUSION

Our first attempt at guitar model identification shows that guitar models do have distinct timbral qualities allowing a trained classifier to distinguish between different guitars. This is shown through the results of the experiments identifying guitar models recorded in the same room. The experiments performed show that we are able to capture enough of the timbral signature of a guitar to successfully identify guitar models when recorded in the same room. Interestingly we are able to achieve classification accuracy above 50% for both recordings configurations B and C. While these two configurations are quite different in timbral qualities, one being a chord progression and the other being a finger picking pattern, the training models are able to capture some aspect of timbre that is consistent between recordings. We are are also able to classify guitar A12 with over 50% accuracy, indicating there is some consistency in timbral signature of different guitars of the same model.

With this work we've also established a baseline for future work into guitar model classification. Affording researchers the ability perform more specific experiments to improve classification models, including the features and classifier used. We have also shown that the room the instruments were recorded in affects identification capabilities. This is a similar problem that researchers in speech recognition have come across as shown in [23].

A. Future Work

We hypothesize that the low recognition results of classifying recordings from different rooms is an effect of reverberation in the rooms. To overcome these effects we plan to implement dereverberation techniques as used in speech recognition

[23]. We would also like to explore why test sets generated from recording configurations A and D have the worst results; possibly through timbral analysis using using clustering or MDS techniques as in [10] and [11]. Gaining an understanding as to why these configurations are not classified as well as B and C, may help us understand what timbral characteristics are most important to guitar model identification, leading us to better recording configurations for training the classification models. This analysis may lead us to improve our classification results by tuning each component model more effectively. Currently we train each model with the same classifier and feature sets. With more experimentation and analysis we may be able to determine better combinations of feature sets and classifiers for each component model. Our first work into this will be to tune the Acoustic Electric classifier. As previously discussed our confusion matrix show that a high percentage of electric guitar segments are misclassified as acoustic guitars.

Acknowledgment

The authors would like to thank Mario Milat for his assistance selecting, playing and recording each of the guitars to create our dataset. We would also like to thank Harry Liang for his support and contribution of ideas to this project.

References

[1] J. Marques and P. J. Moreno, “A study of musical instrument classification using gaussian mixture models and support vector machines,” Compaq Corporation, Cambridge Research laboratory, Tech. Rep., 1999.

[2] G. Agostini, M. Longari, and E. Pollastri, “Musical instrument timbres classification with spectral features,” *EURASIP J. Appl. Signal Process.*, vol. 2003, pp. 5–14, Jan. 2003. [Online]. Available: http://dx.doi.org/10.1155/S1110865703210118

[3] S. Essid, G. Richard, and B. David, “Musical instrument recognition by pairwise classification strategies,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 1401–1412, July 2006.

[4] J. Barbedo and G. Tzanetakis, “Musical instrument classification using individual partials,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 111–122, Jan 2011.

[5] J. Eggink and G. J. Brown, “Instrument recognition in accompanied sonatas and concertos,” in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP ’04). IEEE International Conference on*, vol. 4, May 2004, pp. iv–217–iv–220 vol.4.

[6] J. C. Brown, O. Houix, and S. McAdams, “Feature dependence in the automatic identification of musical woodwind instruments,” *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1064–1072, 2001.

[7] D. Fragoulis, C. Papaodysseus, M. Exarhos, G. Roussopoulos, T. Panagopoulos, and D. Kamarotos, “Automated classification of piano-guitar notes,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 3, pp. 1040–1050, May 2006.

[8] J. M. Grey, “Multidimensional perceptual scaling of musical timbres,” *The Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1270–1277, 1977.

[9] P. Iverson and C. L. Krumhansl, “Isolating the dynamic attributes of musical timbrea),” *The Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2595–2603, 1993.

[10] P. Cosi, G. D. Poli, and P. Prandoni, “Timbre characterization with mel-cepstrum and neural nets,” in *Proceedings of the 1994 International Computer Music Conference*, 1994, pp. 42–45.

[11] G. De Poli and P. Prandoni, “Sonological models for timbre characterization,” *Journal of New Music Research*, vol. 26, no. 2, pp. 170–197, 1997.

[12] A. Eronen and A. Klapuri, “Musical instrument recognition using cepstral coefficients and temporal features,” in *Acoustics, Speech, and Signal Processing, 2000. ICASSP ’00. Proceedings. 2000 IEEE International Conference on*, vol. 2, 2000, pp. II753–II756 vol.2.

[13] A. Eronen, “Comparison of features for musical instrument recognition,” in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001, pp. 19–22.

[14] G. White and R. Neely, “Speech recognition experiments with linear predication, bandpass filtering, and dynamic programming,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 2, pp. 183–188, Apr 1976.

[15] I. Potamifis, N. Fakotakis, and G. Kokkinakis, “Improving the robustness of noisy mfcc features using minimal recurrent neural networks,” in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, vol. 5, 2000, pp. 271–276 vol.5.

[16] B. David and G. Richard, “Efficient musical instrument recognition on solo performance music using basic features,” in *Audio Engineering Society Conference: 25th International Conference: Metadata for Audio*, Jun 2004. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12807

[17] G. Tzanetakis, “Marsyas-0.2: a case study in implementing music information retrieval systems.”

[18] S. Keerthi, S. Shevade, C. Bhattacharyya, and K. Murthy, “Improvements to platt’s smo algorithm for svm classifier design,” *Neural Computation*, vol. 13, no. 3, pp. 637–649, 2001.

[19] G. H. John and P. Langley, “Estimating continuous distributions in bayesian classifiers,” in *Eleventh Conference on Uncertainty in Artificial Intelligence*. San Mateo: Morgan Kaufmann, 1995, pp. 338–345.

[20] D. Aha and D. Kibler, “Instance-based learning algorithms,” *Machine Learning*, vol. 6, pp. 37–66, 1991.

[21] K. D. Martin and Y. E. Kim, “Musical instrument identification: A patternrecognition approach,” *The Journal of the Acoustical Society of America*, vol. 104, no. 3, pp. 1768–1768, 1998.

[22] J. C. Brown, “Computer identification of musical instruments using pattern recognition with cepstral coefficients as features,” *The Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1933–1941, 1999.

[23] B. W. Gillespie, H. Malvar, and D. Florencio, “Speech dereverberation via maximum-kurtosis subband adaptive filtering,” in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP ’01). 2001 IEEE International Conference on*, vol. 6, 2001, pp. 3701–3704 vol.6.