The Significance of the Non-Harmonic "Noise" Versus the Harmonic Series for Musical Instrument Recognition

Arie Livshin **Xavier Rodet**

IRCAM Centre Pompidou
1 place Stravinsky, Paris 75004, FRANCE
arie.livshin,xavier.rodet@ircam.fr

Abstract

Sound produced by Musical instruments with definite pitch consists of the Harmonic Series and the non-harmonic Residual. It is common to treat the Harmonic Series as the main characteristic of the timbre of pitched musical instruments. But does the Harmonic Series indeed contain the complete information required for discriminating among different musical instruments? Could the non-harmonic Residual, the "noise", be used all by itself for instrument recognition? The paper begins by performing musical instrument recognition with an extensive sound collection using a large set of feature descriptors, achieving a high instrument recognition rate. Next, using Additive Analysis/Synthesis, each sound sample is resynthesized using solely its Harmonic Series. These "Harmonic" samples are then subtracted from the original samples to retrieve the non-harmonic Residuals. Instrument recognition is performed on the resynthesized and the "Residual" sound sets. The paper shows that the Harmonic Series by itself is indeed enough for achieving a high instrument recognition rate; however, the non-harmonic Residuals by themselves can also be used for distinguishing among musical instruments, although with lesser success. Using feature selection, the best 10 feature descriptors for instrument recognition out of our extensive feature set are presented for the Original, Harmonic and Residual sound sets.

Keywords: instrument recognition, musical instruments, residual, noise, harmonic series, pitch

1. Introduction

Musical instruments with definite pitch ("pitched instruments") are usually based on a periodic oscillator such as a string or a column of air with non-linear excitation. In consequence, their sound is mostly composed of a Harmonic Series of sinusoidal partials, i.e. frequencies which are integer multiples of the fundamental frequency (*f0*). While the relation between the energy levels of the different harmonics is widely considered as the main characteristic of pitched instruments' timbre (e.g. [1]), if we subtract this Harmonic Series from the original sound there is a non-harmonic Residual left. This Residual is far from being 'white noise'; it is heavily filtered by the nature of the instrument itself as well as the playing technique, and may contain inharmonic sinusoidal partials as well as non-sinusoidal 'noise', such as the breathing sounds in the flute or the scraping noises in the guitar.

Does the Harmonic Series indeed encapsulate all the distinguishing information of the sounds of pitched musical instruments? If so, about the same instrument recognition rate should be achieved by using only the Harmonic Series as by using all the information in the signal, with the same feature descriptor set used for classification. This is a practical question for the field of instrument recognition; when performing instrument recognition in multi-instrumental, polyphonic music, it is difficult as well as computationally expensive to perform full source separation [2] and restore the original sounds out of the polyphonic mixture in order to recognize each source separately. On the other hand, estimating the Harmonic Series of the different notes in the mixture is a relatively easier task [3]. For example, in [4] Harmonic Series estimation is used for performing "Source Reduction", reducing the volume of all instruments except one and then recognizing it. In [1], instrument recognition is performed using only features based on the Harmonic Series, estimated from pre-given notes. There is also research attempting to perform instrument recognition by recognizing directly instrument mixtures, instead of trying to separate them into individual instruments, see for example [5].

Another interesting question comes from the opposite direction: is the non-harmonic Residual, the "noise" a musical instrument produces, so distinct as to allow distinguishing between different instrument types, e.g. can we actually distinguish between different wind instruments just by the sound of their airflow hiss?

In order to answer these questions, the paper explores how instrument recognition rates using signals resynthesized solely from the Harmonic Series of the sound, and signals containing solely the non-harmonic

Residuals, compare with the recognition rates when using the complete signals. In order to perform this comparison as directly as possible, the first step is to achieve a high instrument recognition rate. This is accomplished here by computing an extensive set of feature descriptors on a large and diverse set of pitched musical instrument sound samples, reducing the feature dimensions with Linear Discriminant Analysis (LDA) and then classifying the sounds with K-nearest neighbours (KNN).

Next, the Harmonic Series of each sample in the sound set is estimated, including the *f0*s, harmonic partials and corresponding energy levels, and using Additive Synthesis all the signals are resynthesized using only their Harmonic Series, thus creating synthesized 'images' of the original signals which lack any non-harmonic information. These resynthesized sounds will be referred to in the paper as "Harmonic" signals, while the original sounds from the sound set will be called, the "Original" signals. As the phase information of the Original signals is kept in the Harmonic signals, by subtracting the Harmonic signals from the Original signals we remain with the non-harmonic, "noisy", part of the signals, referred to shortly as the "Residuals".

After that, the same set of feature descriptors is computed on each sample group: the Original, Harmonic and Residual Signals. These three groups are then divided separately into training and test sets and instrument recognition is performed on each group independently. The instrument recognition results are presented and compared in Section 7.

Using the Correlation-based Feature Selection (CFS) algorithm with a greedy stepwise forward search method, the 10 most important feature descriptors for each of the three groups of samples are estimated and presented.

2. Original Sound Set

The sound set consists of 5006 samples of single notes of 10 "musical instruments": bassoon, clarinet, flute, trombone, trumpet, contrabass, contrabass pizzicato, violin, violin pizzicato and piano. As the violin and bass pizzicato sounds are very different from the bowed sounds they are treated here as separate instruments.

The sound samples were collected from 13 different commercial and research sound databases, all "well recorded" and practically lacking noise. The databases contain sounds recorded in different recording environments, using different individual instruments (e.g. using different violins in each sound database). The sound set spans the entire pitch range of each of the 10 instrument types and includes vibrato and non-vibrato sounds where applicable.

The collection of all the samples of a specific instrument taken from a single database (e.g. all the violin samples from database #1), is referred to in the paper as an "instrument Instance". The total number of instrument Instances in the sound set is 77.

All the sounds are sampled in 44 KHz, 16 bit, mono.

3. Harmonic Sounds and Residuals

Additive analysis/synthesis is based on Fourier's theorem, which states that any physical function that varies periodically with time with a frequency *f* can be expressed as a superposition of sinusoidal components of frequencies: *f, 2f, 3f, 4f,* etc. Additive synthesis applies this theorem to the synthesis of sound [6]. For a review of supplementary Additive Synthesis techniques see [7].

In order to separate the sound samples into their harmonic and non-harmonic components, the samples are analyzed and then selectively resynthesized using the Additive analysis/synthesis program, "Additive" [8], which considers also inharmonic deviations of the partials (e.g. found in the piano sounds). Very precise Additive analysis was performed by supplying the Additive program with specifically tailored parameters for each sound sample using its note name and octave, known in advance, for estimating its *f0*. For example, the Additive analysis/synthesis window size was set to 4*(1/*f0*), FFT size to 4*nextpow2[1](sampleRate * windowSize), etc.

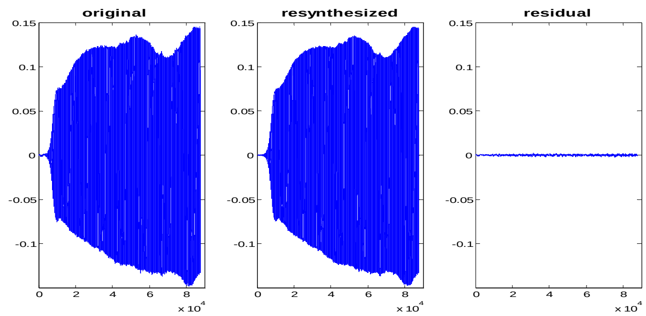

Figure 1. Left to right: original Clarinet sample (A3), the sample resynthesized from the Harmonic Series, the Residual (subtraction).

Figure 1 shows an example of an Original clarinet sample (the note A3), the sound we resynthesized from the Harmonic Series of the Original, and the non-harmonic Residual. We can see that the Original and Harmonic sound envelopes are similar and that the Residual energy, resulting from subtracting the resynthesized Harmonic sound from the Original, is comparatively very low.

While the sounds resynthesized from the Harmonic Series sound very similar to the Original samples, the non-harmonic Residuals sound very differently from them while sounding quite similar to each other for the same instrument. For example, the clarinet Residual of the note A3 sounds like a steady airflow while the trombone Residual of the same pitch sounds mellower and with

[1] nextpow2(N) is the first P such that 2^P >= abs(N)

addition of “static-electricity” crackle. The bass pizzicato Residual of A3 sounds like a wooden barrel being hit with a hammer, while the Residual of the violin pizzicato of exactly the same note sounds much higher “pitched” due to the considerably smaller size of its wooden resonator, and includes some tremolo. To learn how the physical structure of musical instruments shapes the sound, see [9]. Note that the Attacks are far from being the only parts of the Residuals influencing the descriptors; The Sustained parts of the Residuals of the clarinet, flute, trombone and trumpet contain energy levels as high as or higher than their Attack Transients.

4. Feature Descriptors

The same feature set[2] is computed on the Original samples, the Harmonic samples and the Residuals.

In order to encapsulate various characteristics of the signals, the feature set is quite large and includes 62 different feature types. Many of these features include several variations using different parameter types, resulting in a total of 513 different feature descriptor “flavors”. For example, Spectral Kurtosis “flavors” include Kurtosis computed on the linear spectrum, the log-spectrum, the harmonics envelope, etc. After computation, the feature descriptors are normalized to the range of [0 - 1] using Min-Max Normalization. Except the features computed on the whole signal, most of the features are computed on the Short-Time Fourier Transform (STFT) of the signal, using a sliding frame of 60 ms with a 66% overlap. For each sample, the average and standard deviation of these frames are used as feature descriptors.

The different feature types are:

4.1 Temporal Features

Features computed on the whole signal (without division into frames), such as Log Attack Time, Temporal Decrease, Effective Duration, etc.

4.2 Energy Features

Features referring to the energy content of the signal, like Total Energy, Harmonic Energy, Noise-Part Energy, etc.

4.3 Spectral Features

Features computed from the Short Time Fourier Transform (STFT) of the signal, including the Spectral Centroid, Spectral Spread, Spectral Skewness, etc.

4.4 Harmonic Features

Features computed from the Sinusoidal Harmonic modeling of the signal, like *f0*, Inharmonicity, Odd to Even Ratio, etc.

[2] The feature computation routines were written by Geoffroy Peeters of IRCAM. Full feature list can be found in [10].

4.5 Perceptual Features

Features computed using a model of the human hearing process, including Mel Frequency Cepstral Coefficients (MFCC), Loudness, Sharpness, etc.

5. Feature Selection

In order to provide the 10 best features out of our extensive feature set for each group of samples (the Original samples, the Harmonic samples and the Residuals) the Correlation-based Feature Selection (CFS) evaluator is used with a greedy stepwise forward search method. The entropy-based CFS algorithm scores and ranks the “worth” of subsets of features by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. As the feature space is very large and checking all the feature combinations is not practical, CFS starts with an empty set and adds features using a stepwise forward search method, searching the space of feature subsets by greedy hillclimbing augmented with a backtracking facility. For further reading on CFS see [11]. In this paper, we use the WEKA data-mining software [12] implementation of the CFS algorithm.

6. Classification and Evaluation

Instrument recognition is performed on the Original, Resynthesized and Residual sets of samples separately.

6.1 Minus-1 Instance Evaluation

In order to get meaningful instrument recognition results it is necessary not to use sounds recorded by the same instrument and the same recording conditions both in the learning and test sets [13]. For this purpose, we introduce the ‘Minus-1 Instance’ cross-validation evaluation method: each instrument Instance is removed in its turn from the sound set[3] and classified by all the remaining samples. The recognition rate is computed per instrument type and is the average of the grades of its Instances.

6.2 Classification

Each classification phase of the Minus-1 Instance Evaluation begins by computing a Linear Discriminant Analysis (LDA) transformation matrix using the learning set. LDA [14] reduces the dimensionality of data with C classes down to C-1 dimensions while maximizing the distance between the means of the different classes and minimizing the variance inside each class (the Fisher criterion). After dimension reduction, the test set is

[3] Reminder: our sound set is joined from samples originating from 13 different sound databases. At each classification step, all the samples of one instrument from one of these databases are removed.

classified by the learning set using the K-Nearest-Neighbors (KNN) algorithm. K values in the range of [1 - 80] are tested at each classification phase. After the Minus-1 Instance evaluation process completes and all the Instances are classified, the best K for the whole classification process is reported.

7. Results

7.1 Instrument Recognition

The confusion matrices in this section show the Minus-1 Instance recognition rates for the Original samples, the Harmonic samples and the Residuals. These matrices show the percentage of samples (rounded to integers) of the instruments in the first column which were classified as the instruments in the first row. For example in Table 1, 8% of the clarinet samples are misclassified as flute. The instrument abbreviations are: bsn = Bassoon, cl = Clarinet, fl = Flute, tbn = Trombone, tr = Trumpet, cb = Contrabass, cbp = Contrabass Pizzicato, vl = Violin, vlp = Violin Pizzicato, pno = Piano.

7.1.1 Original Samples

Table 1. Confusion matrix of the Original samples

	bsn	**cl**	**fl**	**tbn**	**tr**	**cb**	**cbp**	**vl**	**vlp**	**pno**
bsn	**95**	0	1	3	0	0	0	0	0	0
cl	0	**89**	8	0	1	1	0	0	0	0
fl	0	4	**94**	0	0	0	0	1	0	0
tbn	2	0	0	**94**	2	0	0	1	0	0
tr	0	2	2	1	**95**	0	0	0	0	0
cb	0	0	0	0	0	**98**	0	1	0	0
cbp	0	0	0	0	0	0	**98**	0	0	2
vl	0	0	2	0	0	0	0	**97**	0	0
vlp	0	0	0	0	0	0	1	0	**94**	5
pno	1	0	0	1	0	2	0	0	1	**94**

The average Minus-1 Instance recognition rate per instrument for the Original samples is 94.89% (using K=18 with KNN). It is rather hard to compare recognition rates with other papers as each paper attempts to recognize its own instrument set, uses different sound databases and different evaluation techniques. In addition, most papers on instrument recognition of separate tones are unfortunately using sounds from the same instrument Instances both in the learning and test sets, a fact which raises a strong doubt regarding the applicability of their results, which are often unrealistically high [13]. Even so, while our results are obtained by Minus-1 Instance evaluation, they are still higher or comparable to most instrument recognition rates reported by papers on instrument recognition of separate tones, regardless of their evaluation techniques. In [13] for example, an average Minus-1 DB recognition rate of 83.17% for 7 instruments is achieved. It is interesting to note, that the main difference between the classification performed in this section and the one we used in [13] is that our current sound set is much larger and more diverse. This exemplifies well an intuitive claim from [13], which states that enriching a learning set with sound samples from different databases improves its generalization power.

7.1.2 Harmonic Samples

Table 2. Confusion matrix of the Harmonic samples

	bsn	**cl**	**fl**	**tbn**	**tr**	**cb**	**cbp**	**vl**	**vlp**	**pno**
bsn	**93**	2	0	1	1	2	0	0	0	0
cl	2	**86**	7	0	5	1	0	1	0	0
fl	0	7	**89**	0	1	0	0	3	0	0
tbn	5	0	0	**84**	8	2	0	0	1	0
tr	1	5	6	3	**84**	1	0	1	0	0
cb	1	0	0	0	0	**96**	0	2	0	0
cbp	0	0	0	0	0	0	**98**	0	1	1
vl	0	1	3	0	0	2	0	**93**	0	0
vlp	0	0	0	0	0	0	3	0	**90**	6
pno	1	0	0	0	0	1	2	0	3	**92**

The average Minus-1 Instance recognition rate per instrument for the resynthesized samples is 90.53% (using K=4 with KNN). This recognition rate is only 4.36% lower than the rate achieved using the Original samples, and is still quite high. This rate shows that the information in the Harmonic Series of the signal is quite enough for achieving a high average instrument recognition rate which is rather close to the rate obtained using the complete signals. Comparing the confusion matrices of the Harmonic samples (Table 2) to the Originals (Table 1), we can see that the recognition rate of all the instruments has worsened somewhat, which consistently indicates that some instrument-discriminating information was lost. The most noticeable declines are the trumpet (-11.25%) and the trombone (-9.42%).

7.1.3 The Residuals

Table 3. Confusion matrix of the Residuals

	bsn	**cl**	**fl**	**tbn**	**tr**	**cb**	**cbp**	**vl**	**vlp**	**pno**
bsn	**89**	2	2	1	0	0	0	1	0	4
cl	1	**53**	24	3	9	3	0	4	0	4
fl	0	23	**65**	1	0	2	0	9	0	0
tbn	2	4	1	**77**	5	0	0	0	1	10
tr	2	29	12	0	**56**	1	0	0	0	0
cb	0	2	1	0	0	**97**	0	0	0	1
cbp	0	0	0	0	0	0	**95**	0	0	5
vl	0	9	13	0	0	0	0	**76**	0	1
vlp	1	1	1	0	0	1	0	1	**86**	8
pno	3	0	0	1	0	3	2	2	3	**85**

The average Minus-1 Instance recognition rate for the Residuals is 77.94% (using K=21 with KNN), which is 16.95% lower than the rate achieved with the Original samples. While this is a considerable difference, these results do indicate, perhaps surprisingly, that the

Residuals by themselves (yes, these “airflow” and “click” sounds) contain considerable distinguishing instrument information. As this experiment did not involve any descriptors “tailored” specifically for the Residuals, it seems reasonable to expect that the recognition rate could be improved further.

The instruments with recognition rates reduced by more than 25% compared with the Original samples are the clarinet (-36.56%), flute (-29.24%), confused mainly with each other, and the trumpet (-38.44%), also confused mostly with the clarinet and flute.

7.2 Best 10 Feature Descriptors

Using CFS with a greedy stepwise forward search method, the best 10 feature descriptors were selected for each of the three sample groups out of the total 513 different feature descriptors in our feature set.

Table 4. 10 best features for the Original, Harmonic and Residual sample groups, selected using CFS.

Feature Type	Descriptor Flavor	ST	O	H	R
Rel. Specific Loudness	2’nd Mel-Band	m	1	1	1
Temporal Increase			2	2	X
Spec. Kurtosis	log freq., norm. db ampl.	m	3	X	3
Temporal Centroid			4	3	2
MFCC	2’nd coefficient	m	5	X	X
Delta-Delta MFCC	1’st	m	6	X	9
Spec. Spread	lin. freq., norm. db ampl.	m	7	4	X
Temporal Decrease			8	6	X
Roughness	mean (ERBs)		9	X	X
Bark-Band Tristimulus	lin. ampl, bands(2+3+4)/sum(all)	m	10	10	X
Bark-Band Tristimulus	norm. db ampl., band(1)/sum(all)	m	X	5	X
Spec. Skewness	lin. freq., norm. db ampl.	m	X	7	X
Harmonic Spec. Roll-Off		m	X	8	X
Fluctuation Strength	7’th ERB	m	X	9	X
Spec. Variation	norm. db ampl.	s	X	X	4
MFCC	4’th coefficient	m	X	X	5
Perceptual Spec. Kurtosis	orig. bands, orig. ampl.	s	X	X	6
Fluctuation Strength	mean (ERBs)		X	X	7
Bark-Band Tristimulus	quad. freq., sum(5:end)/sum(all)	m	X	X	8
Spec. Kurtosis	lin. freq., norm. db ampl.	m	X	X	10

The “Feature Type” column in Table 4 shows the feature type, while the “Descriptor Flavor” column shows the parameter types used with this feature. “Feature Type” column abbreviations: Rel. = Relative. spec. = spectral. “Descriptor Flavor” column abbreviations: freq. = frequency scale, ampl. = amplitudes, lin. = linear, orig. = original, quad. = quadratic, norm. = normalized. For a comprehensive description of all these features, see [10].

Most features are computed on each STFT frame of the signal separately and then either the mean (‘m’) or the standard deviation (‘s’) of these frames is used. For such features, the “Frames” column specifies which of these statistics was used. The “O”, “S” and “R” columns indicate the Original sample group, the Harmonic samples (Synthesized) and the Residuals, and show which feature descriptors were selected for these sample groups and in which order of importance, from 1 to 10. An X indicates that a feature was not selected.

The recognition rates using only these sets of 10 selected feature descriptors are 71.18% for the Original samples, 71.43% for the Harmonic samples and 64.48% for the Residuals.

The Original samples “share” 6 descriptors with the Harmonic samples and 4 with the Residuals out of the selected 10. The Harmonic samples and the Residuals share among themselves only 2 descriptors: the Temporal Centroid and the Relative Specific Loudness in the 2’nd Mel-Band; these descriptors are also shared with the Original samples. It is interesting to note that the Relative Specific Loudness in the 2’nd Mel-Band is the most prominent feature descriptor (#1) selected for all three groups, and thus seems to be very appropriate for instrument recognition in general.

Although Table 2 shows that the Harmonic signals contain most of the distinguishing instrument information (resulting in a recognition loss of only 4.36% compared with the Original signals), Table 4 shows that besides this small decrease in average recognition rate, removing the non-harmonic residuals has also caused a somewhat different set of features to be selected by the feature selection algorithm. This indicates that the non-harmonic Residuals present in the Original signals do influence the instrument recognition process.

8. Conclusions

The paper shows that using only information present in the Harmonic Series of the signal is enough for achieving a high average musical instrument recognition rate – 90.53% for 10 instruments using Minus-1 Instance evaluation. This is only 4.36% less than the recognition rate obtained by using the complete, Original signals.

On the other hand, Table 3 shows that there is a lot of distinguishing instrument information present in the non-harmonic Residuals which by themselves produced an average instrument recognition rate of 77.94%. It was also shown that the information present in the non-harmonic Residuals is not completely redundant to the information present in the Harmonic Series; Table 2 shows that although the average recognition rate of the Harmonic signals is high, some of the instruments have suffered noticeably from removing the non-harmonic Residuals, especially the trumpet and trombone. In addition, Table 4 shows that the 10 best feature descriptors selected for the Original sample set differ from the ones selected for the Harmonic samples. These results show that the sound of pitched musical instruments should not be treated as containing only the Harmonic Series, although most of the energy and distinguishing instrument information of the signal is indeed present in the Harmonic Series.

9. Future Work

It was shown that using only the harmonic series does not considerably lower the average instrument recognition rate although some instruments "suffer" more than others. This means that instrument recognition in polyphonic, multi-instrumental music could indeed be performed with rather high results without performing full source-separation; Using multiple *f0* estimation algorithms (such as [2]), estimated harmonic partials could be used solely to classify musical instruments without losing too much distinguishing information.

It might be possible to increase the instrument recognition rate of the Residuals by specifically tailoring special feature descriptors for them. Instrument recognition of pitched instruments could then be improved by splitting the classified sounds into harmonic and non-harmonic components (when applicable) and computing special feature descriptors on the Residuals in addition to the feature descriptors computed on the original signal. The splitting of the signal makes it easier to deal with the non-harmonic Residuals, due to their relatively low energy.

10. Acknowledgments

This work is partly supported by the "ACI Masse de données" project "Music Discover".
Thanks to Geoffroy Peeters for using his feature computation routines.

References

[1] J. Eggink and G. J. Brown. "Instrument recognition in accompanied sonatas and concertos," in *ICASSP 2004 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing Proc.*, 2004, pp. 217-220.

[2] E. Vincent and X. Rodet. "Instrument identification in solo and ensemble music using Independent Subspace Analysis," in *ISMIR 2004 Fifth Int. Conf. on Music Inf. Retr. Proc*., 2004.

[3] C. Yeh, A. Röbel and X. Rodet. "Multiple fundamental frequency estimation of polyphonic music signals," in *ICASSP 2005 IEEE Int. Conf. on Acoustics, Speech and Signal Processing Proc.*, 2005.

[4] A. Livshin and X. Rodet. "Musical Instrument Identification in Continuous Recordings," in *DAFx 2004 7'th international conference on Digital Audio Effects Proc.*, 2004, pp. 222-227.

[5] S. Essid, G. Richard and B. David. "Instrument recognition in polyphonic music based on automatic taxonomies," *IEEE Transactions on Speech and Audio Processing*, Jan. 2006.

[6] J. C. Risset. "Computer Music Experiments, 1964-…," *Computer Music Journal*, vol. 9, no. 1, pp. 11-18, 1985.

[7] X. Serra and J. O. Smith. "Spectral Modeling Synthesis: A Sound Analysis Synthesis System Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12-24, 1990.

[8] "IRCAM Real time applications - Additive and HMM," Available:http://www.ircam.fr/58.html?L=1&tx_ircam_pi2%5BshowUid%5D=35&ext=2

[9] N. Fletcher and T. D. Rossing. *The Physics of Musical Instruments, second edition*. Springer, 1998.

[10] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," 2003, Available:http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf

[11] M. A. Hall, *Correlation-based feature selection machine learning*, Ph.D. Thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1998.

[12] I. H. Witten and E. Frank. *Data Mining – Practical Machine Learning Tools and Techniques, second edition*. Morgan Kaufmann Publishers (imprint of Elsevier), 2005.

[13] A. Livshin and X. Rodet. "The Importance of Cross Database Evaluation in Musical Instrument Sound Classification," in *ISMIR 2003 Fourth Int. Conf. on Music Inf. Retr. Proc.*, 2003.

[14] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. New York, NY: Wiley Interscience, 1992.