SCALABILITY, GENERALITY AND TEMPORAL ASPECTS IN AUTOMATIC RECOGNITION OF PREDOMINANT MUSICAL INSTRUMENTS IN POLYPHONIC MUSIC

Ferdinand Fuhrmann, Martín Haro, Perfecto Herrera
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
{ferdinand.fuhrmann,martin.haro,perfecto.herrera}@upf.edu

ABSTRACT

In this paper we present an approach towards the classification of pitched and unpitched instruments in polyphonic audio. In particular, the presented study accounts for three aspects currently lacking in literature: model scalability to polyphonic data, model generalisation in respect to the number of instruments, and incorporation of perceptual information. Therefore, our goal is a unifying recognition framework which enables the extraction of the main instruments' information. The applied methodology consists of training classifiers with audio descriptors, using extensive datasets to model the instruments sufficiently. All data consist of real world music, including categories of 11 pitched and 3 percussive instruments. We designed our descriptors by temporal integration of the raw feature values, which are directly extracted from the polyphonic data. Moreover, to evaluate the applicability of modelling temporal aspects in polyphonic audio, we studied the performance of different encodings of the temporal information. Along with accuracies of 63% and 78% for the pitched and percussive classification task, results show both the importance of temporal encoding as well as strong limitations of modelling it accurately.

1. INTRODUCTION

Instrument recognition is one of the big problems of current research in music information retrieval (MIR). Automatic indexing and retrieval of audio data are basic concepts to efficiently administrate and navigate through big datasets. Providing the information about the instrumentation of audio tracks via an automatic recognition system can highly facilitate these operations. Besides, such a system provides higher-level musical information, which helps to narrow the well-known semantic gap [1].

Computational recognition of musical instruments makes use of the intrinsic properties of, and differences between, each of the target categories. In the case of pitched instruments, where the sound is mostly composed of quasi-harmonic components, these are the amplitudes and frequency positions of the components and their evolution in time. The time-varying spectral envelope, an eminent feature for pitched instrument recognition [2], can be estimated out of them. For percussive instruments, properties such as attack and decay time, or frequency coverage, are properties which allow to distinguish between them [3]. While these specific characteristics can be determined without big problems in the case of a monophonic recording, the problem gets harder in polyphonic audio. Since the co-occurrence of multiple sound sources is producing overlapping frequency components, information extracted from the raw audio is often ambiguous and only partially useful for discriminating between several musical instruments. Without any preprocessing based on source separation, which is still not mature enough, models derived from simplified scenarios seem to imply strong limitations for the use on polyphonic audio. However, we hypothesize that, by providing a well suited dataset, a coarse – but MIR useful – modelling of predominant instruments directly from polyphonic audio is possible.

Below follows a short review of the current state of the art in computational musical instrument recognition. In Sec. 3 we substantiate our work and provide details about the general concepts we used for tackling the problem. Sec. 4 gives insights in the used data, the developed algorithms, and shows the experimental results. In the subsequent discussion we point out capacities as well as limitations of the chosen techniques and, finally, Sec. 6 concludes this article.

2. RELATED WORK

In current literature there exists a great unbalance between the amount of studies dealing with recognition of pitched instruments from polyphonic data and the amount of publications studying the monophonic case. Since the latter is not addressed in this paper, we refer to [4] for a comprehensive overview. Regarding the scarce publications addressing the more complex scenario, Kitahara et al. [5] presented a method to eliminate unreliable feature data caused by source inference for instrument recognition in artificial polyphonic mixtures. Linear discriminative analysis (LDA) was used to enhance features which discriminate best the five categories. The features were extracted from the harmonic structures of the corresponding instruments

and used to train multivariate gaussian prototypes. Additional post-processing was applied by integrating the frame-wise a-posteriori probabilities and incorporating higher-level musical knowledge to get the final classification. In a more recent work, Every [6] manually annotated songs from a commercially available collection according to perceptually dominant instruments along with their corresponding pitches. A great set of audio descriptors was extracted from the raw audio in an unsupervised system, where the feature vectors were clustered and the resulting accuracies were measured. A strategy for tackling the problem at hand from a complete different direction was presented by Essid et al. [7]. Unlike trying to isolate the instruments present in the mixture, the whole audio was classified considering the more frequent combinations of them. Therefore, a suitable taxonomy was automatically generated by clustering a training corpus. Statistical models were built for each of the derived categories and used to classify unseen instances.

Regarding the recognition of percussive events in polyphonic music, work has focused on transcription of most common drum kit sounds (i.e. Bass Drum, Snare Drum and Hi-Hat sounds). For an excellent overview of studies on drum transcription up to 2006 see [3]. In a more recent work Paulus and Klapuri [8] evaluated a system based on Hidden Markov Models (HMM). In addition to standard spectral features, temporal features were derived from sub-band envelopes using 100 ms windows. Slight improvements in transcription accuracy were reported by incorporating this temporal information. Gillet and Richard [9] used source separation as preprocessing to obtain a drum-enhanced signal. A set of features was computed from both original and "enhanced" signals. Classification was derived from previously trained support vector machines (SVM) on an experimental database consisting of 28 songs from the ENST database (see Sec. 4.1 for an overview of this database).

Examining the literature review above we detect three main gaps in which we substantiate our present work. More precisely, we miss the aspect of *polyphonic scalability*, i.e. a detailed research about the application of current methods for instrument recognition to highly polyphonic audio. Second, there does not exist, to our knowledge, a study accounting for *instrument generality*, i.e. presenting a consistent methodology incorporating multiple instruments from different musical styles for tackling this problem. Finally, we see a clear need for incorporating *temporal characteristics* within the recognition process when working with statistical models, as this information is known to be important but often neglected.

3. CONCEPTUAL OVERVIEW

The present study is thought as a first step towards assessing the aforementioned gaps. We propose a methodology using statistical recognition techniques to build independent classification systems for 11 pitched and 3 unpitched instruments. Ground truth obtained from mostly manually created collections is used for training the models, gathered only from real world music. Additionally, we evaluate the importance and modelling accuracy of temporal aspects by comparing systems using different encodings of temporal information. What follows is a more detailed description of the concepts addressed within this work.

Polyphonic scalability. In this paper we are taking an approach of learning the time-frequency characteristics of musical instruments directly from polyphonic data. Our aim is to label a given audio excerpt with the name(s) of the most salient instrument(s). There exist some evidence that the performance of a recognition system is improved when the polyphonic context is incorporated into the training process [10]. As we focus on the application of a recognition algorithm on commercially available music, we introduce the least simplified conditions and work directly with real world data, all containing predominant instruments plus accompaniment.

Instrument generality. We included the recognition of pitched as well as unpitched (percussive) instruments in our study. As they imply obvious differences in their sound characteristics, both groups have to be treated in a slightly different way for computational processing. Percussive instruments produce a high energetic, impulsive sound and carry the main information in a relative short time interval (typically between 100 and 200 ms), whereas pitched instruments tend to have a quasi-harmonic and continuous tone ranging from very short to medium long durations (several seconds). Moreover, percussive sounds produce a spectrum in which their energy is scattered among the frequency bins, whereas pitched instruments have a frequency representation with peaks at quasi-integer multiples of their fundamental. Furthermore, a clear pitch dependency of the spectrum can be observed with pitched instruments unlike the more fixed spectral patterns of drum sounds.

Pitched Instruments: We consider an instrument to be "pitched" if it is able to produce a continuous, quasi-harmonic sound. Ten pitched instruments (Cello, Clarinet, Flute, acoustic and electric Guitar, Hammond Organ, Piano, Saxophone, Trumpet and Violin) are used in this study, being a good representation for most of the possible instrumentations in real world music of Western culture. We also include the human singing voice as an extra category in the corpus, as it can bee seen as frequently used pitched instrument in pop and rock music.

Unpitched Instruments: Due to the importance of the drum kit in Western popular music we decided to concentrate our research efforts on this particular set of percussive instruments. Likewise, because of the number of available instances and the musical relevance of each instrument within the drum kit, we work with the following, most common in literature, instrument classes: Bass Drum (BD), Snare Drum (SD) and Hi-Hat (HH).

Temporal characteristics. We incorporate temporal information in our statistical modelling process. According to experimental findings in literature, the human auditory system uses temporal aspects as an important cue for the recognition of musical instruments [2], but this informa-

tion is often neglected in related studies. Together with quasi static properties of the sound, its evolution in time shapes the basics of human timbre perception. Therefore we directly compare different encodings of the temporal information in our statistical models. Doing that, we do not only examine the applicability of these aspects to computational musical instrument recognition. We also study how far they can be modelled directly from the polyphonic audio.

In [11] the modelling of temporal aspects was already analyzed by comparing the performance of a monophonic to a polyphonic similarity task. Timbre similarity was evaluated by both static systems, ignoring temporal aspects, and algorithms incorporating temporal behaviour. The used data consisted of isolated sound samples for the monophonic similarity task, and one song from The Beatles, segmented into its individual notes, for the polyphonic case. The authors concluded their work by stating that the frame-based analysis of polyphonic audio is not suited for modelling any temporal related properties. Moreover, as the performance of the dynamic systems was superior for the monophonic analysis and the same amount inferior for the polyphonic scenario, they identified the polyphony itself to be the root of all evil. However, we try to tackle this problem of polyphony by using big, diverse datasets and show that there still remain temporal aspects which can be modelled, if not for similarity retrieval, at least for sound source recognition.

4. METHOD

4.1 Data

A key concept for a successful modelling of musical instruments from polyphonic audio is the quality and the representativeness of the used data. We used two public available datasets to form the corpus for the recognition of percussive instruments, and developed our own collection for the pitched instrument identification task. Therefore we manually gathered sound samples from the three considered super-genres of Western music (jazz, classical and pop/rock), all extracted from commercial available recordings.

The objective for the creation of the dataset for the pitched instruments was to assemble excerpts of polyphonic audio in which the target instrument is playing continuously and is easily audible for a human listener. Each audio excerpt was then labelled with its predominating instrument (double-checked by two human experts), thus assigning more than one instrument to an audio excerpt was not allowed. After all, a corpus containing about 2,500 audio files was created, each one taken from a different recording. We tried to equally distribute the data among the three above-mentioned super-genres in order to cover most musical styles and combinations of instruments.

In the case of percussive instruments we used two publicly available collections with proper annotations of percussive events, namely the ENST-Drums database [12] and the MAMI database [13]. The first one is the largest publicly available drum database which provides "wet" and "dry" (see [12] for detailed information) drum tracks, as well as the respective accompaniment tracks. We decided to work with the "wet" drums and their accompaniment. From the obtained collection of 64 songs we randomly selected 30 second excerpts of every song and its labels. The MAMI database is a collection of 52 annotated music fragments extracted from commercial audio recordings. We managed to gather 48 songs and aligned them with the provided annotations. Finally, we mixed the ENST and the MAMI databases in order to have a representative database for training purposes. Thus we obtained a large set of polyphonic music excerpts adding up a total of 112 songs labelled with three, possibly concurrent, tags.

4.2 Algorithm Processing

Our approach towards assessing the information encoded by the different instruments and developing suitable models is based on classical pattern recognition techniques. First, we extract segments from the audio file containing the target instrument. For the drum recognition algorithm we generate excerpts based on onset detection: either we take a segment starting from the onset and lasting for 150 ms or, if the next onset falls within the following 150 ms, we take the inter-onset-interval. In the end, we include every so-generated excerpt in the dataset. For the pitched instruments we randomly extract a maximum of four 2.5 s long segments from each audio file. This grants a big amount of variability in the polyphonic background, which accompanies the main instrument. The length of 2.5 s was empirically determined and showed superior performance over shorter durations, whereas no significant improvement could be observed by using longer excerpts.

These segments are then framed with a fixed framesize of 46 ms and hopsize of 12 ms using a Blackman-Harris windowing function and audio features are extracted for every frame. We use a big amount of spectral, cepstral, and tonal features, all of them are well known audio descriptors and will not be discussed here. For a comprehensive overview of standard audio features we refer the interested reader to [14].

The frame-wise extraction results in a time series of feature vectors, consisting of the raw feature values. This two dimensional representation (features versus frames) is further processed by describing the evolution in time of each audio feature. We compute standard statistical measures like mean, variance from both the actual and the delta values, as well as more specific quantities accounting for the temporal information. The full set of the applied functions together with a short description is listed in Table 1. Finally, we derive one vector with a dimension of 2,023 representing the audio content of the extracted segment.

To decrease the complexity of the problem we perform feature selection on our data. For our experiments we search for the best subset of descriptors in the feature space, taking their correlation with the respective classes and their intercorrelation inside the subset into account [16]. We apply a 10 fold cross-validated feature selection to return

name	description
mean	mean of the values
var	variance of the values
dmean	mean of the delta values
dvar	variance of the delta values
max-norm-pos	location of the maximum
min-norm-pos	location of the minimum
attack	slope of the attack
decay	slope of the decay
slope	overall slope
t-centroid	temporal centroid of the values
t-skewness	temporal skewness of the values
t-kurtosis	temporal kurtosis of the values

Table 1. Applied functions to describe temporal aspects of the raw feature values. See [15] for details on their implementation.

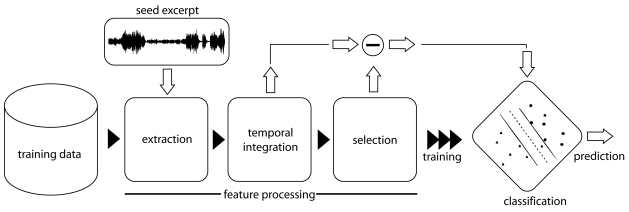

Figure 1. Block diagram of the training and recognition process. Black arrows indicate the training process while white ones show the prediction cycle. Note the decoupled modules of *extraction*, *temporal integration* and *selection* in the feature processing stage.

both a discriminative and compact set of descriptors. This procedure reduces the dimensionality of the vectors by a factor of 20, which significantly lowers the computation time of the following steps.

The feature vectors are then used to train SVMs, powerful classifiers for complex classification tasks. As the SVM is a binary classifier by definition, different strategies for combining the data and training the SVMs were tested to apply them to the multi-class problem. For our drum recognition system we utilized a balanced one–versus–all schema, where one SVM discriminates between the target category and an artificial one, consisting of a mixture of the remaining classes. Hence, each classifier determines the presence or absence of the respective class. In the case of pitched instruments we use a balanced one–versus–one algorithm with pair-wise coupling (PWC) [17], which performed superior than the one–versus–all approach in preceding experiments. Here, the final decision about the class membership is made by combining the output probabilities of all binary SVMs. The so generated models are then used to predict the labels of new data, represented as feature vectors. Fig. 1 shows an overview of the presented algorithm with a detailed view on the feature processing stage.

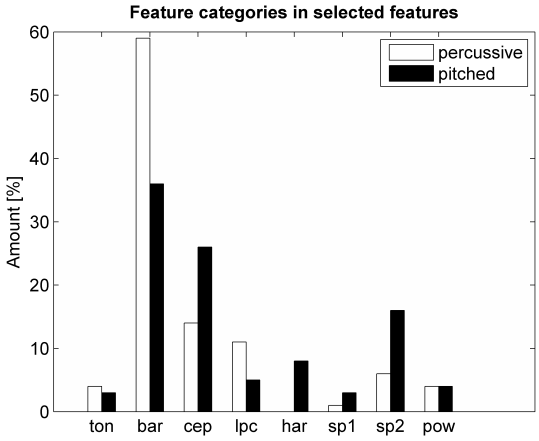

Figure 2. Relative occurrences of different feature categories in the final feature selection, applied to the full set of descriptors. The categories are derived in respect to the acoustic facets the features represent. See text for a detailed description.

4.3 Experiments and Results

First we evaluated the application of our features in the context of the pitched and the unpitched classification tasks. We grouped all selected descriptors in respect to the acoustic facets they represent: 8 categories were derived to evaluate the relative importance of the raw features when extracted from polyphonic data. In particular, the categories included *ton* (HPCPs, pitch salience), *bar* (Barkband energies), *cep* (MFCCs), *lpc* (LPCs), *har* (tristimuli, inharmonicity, odd2even), *sp1* (the four spectral moments), *sp2* (crest, rolloff,...), and *pow* (RMS, 3-bandenergies). Their relative occurrences are shown in Fig. 2. Furthermore, to assess the importance of temporal information encoded in the selected descriptors, we again grouped all of them into three new subsets. According to their modelling of the temporal information we derived the categories μ/σ^2 (only the average and deviation of the values), Δ (coarse encoding of the temporal characteristics in the delta descriptors), and time (detailed modelling of temporal aspects). Fig. 3 shows the results.

For the final evaluation of the recognition systems we split our data into two sets of 90 and 10% of their sizes. 10 fold cross validation was performed on the 90% dataset while the remaining 10% were used as an independent hold-out test set. Performance was measured by the resulting classification accuracy. Furthermore, to evaluate the effectiveness of the descriptors derived by the temporal integration of the raw feature values, we compared the performance of 3 different feature subsets. The first subset consisted of the full set of features, the second contained the average and the variance of both the actual and the delta feature values and subset 3 only included the mean and the variance of the raw values. Hence, we look at different encodings of the temporal information and their application for the recognition process. For all three groups we performed the above described feature selection procedure

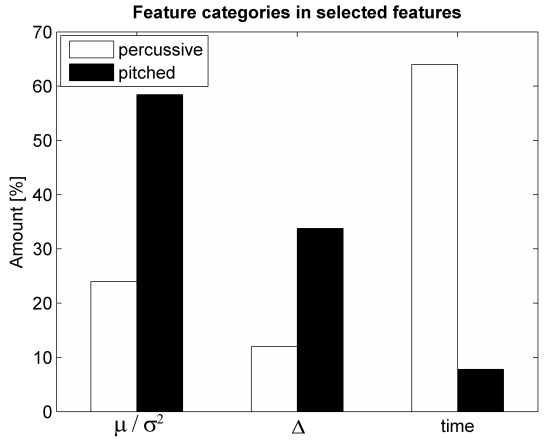

Figure 3. Relative occurrences of different descriptor categories in the final feature selection, applied to the full set of descriptors. The categories represent increasing encodings of the temporal information.

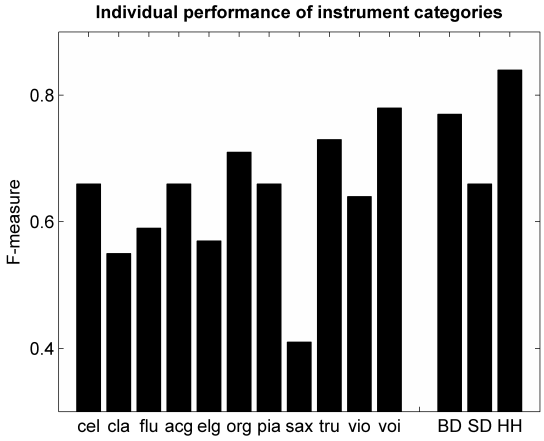

Figure 4. System performance in correctly identifying individual instrument categories. F-measures of the 10 Fold Cross Validation on the $\mu/\sigma^2 + \Delta$ feature subset.

data	full set	μ/σ^2+Δ	μ/σ^2
Pitched	63.1 / 50.3%	63.4 / 50.3%	61.1 / 47.2%
Drums	77.8 / 78.1%	77.6 / 82.6%	74.7 / 78%

Table 2. Performance comparison of different feature subsets with decreasing incorporation of temporal information. Accuracy of 10 fold cross validation (left values) and the hold-out set (right values) is shown (values for drums represent averaged individual accuracies).

before classification. The resulting accuracies can be seen in Table 2. Additionally, Fig. 4 provides details about the system performance in correctly identifying the individual classes on subset 2.

A binomial test [18, p. 37] revealed a significant difference between the cross-validated accuracies in the first two columns of Table 2 and those from the third column, for both pitched and unpitched instruments (p-value of the null hypothesis $\leq 10^{-3}$). Obviously, no statistical significance was found by comparing the first two columns for both recognition algorithms.

5. DISCUSSION

The above presented results show the capacities of the chosen approach as well as some clear limitations. By using a big, well suited dataset covering all relevant musical styles and a selected set of audio descriptors, the algorithm is able to learn the time-frequency characteristics of different musical instruments even from polyphonic data to a certain degree. This indicates that, although the target instrument is partly masked by various accompanying sounds, there still exist information in the audio data which is correlated with the main instrument. Moreover, our selected audio features can be used for extracting this intrinsic information from complex mixtures.

Looking at the results of the grouping experiment presented in Fig. 2 we can infer that both Barkband energies and cepstral features play a major role in discriminating between instrument classes. In particular, in the drum recognition the Barkbands form about 60% of all selected descriptors. This confirms that the spectral energy distribution is an important characteristic of percussive instruments. In the case of pitched instruments, the number of cepstral and spectral descriptors selected indicates the importance of the spectral envelope for recognition.

However, apart from the imbalance of categories (11 vs. 3), the performance differences between the pitched and percussive recognition indicate that the former task is more complex than the latter. This can also be derived from the fact that the characteristics of pitched instruments are more difficult to capture with the current audio features. As the required information is carried in a few frequency bins, a lot of noise due to overlapping components is incorporated into the feature values. Furthermore, percussive sounds generally carry more energy at the same time scale and therefore exhibit a more robust feature extraction.

Examining the performance of the individual instruments (Fig. 4), we can observe that the Snare Drum performs worse in respect to Bass Drum and Hi-Hat. As the latter ones only cover the very low and high frequency regions respectively, the Snare Drum has to compete with several other instruments in the same region, which degrades the systems' performance in correctly labelling Snare Drum sounds. Regarding pitched instruments, the weakness of the saxophone can be explained by its variety inside the class (e.g. Bass, Baritone, Tenor, and Alto), in contrary to other classes (e.g. hammond organ). Interestingly, the singing voice performs best among the pitched instruments, which was not expected. As an additional support of these observations, it is worth mentioning that similar results were obtained by the hold-out test set.

Nevertheless, compared to the human ability to recognize sounds, which is still the measure of all things, the results clearly indicate an inferior performance of our ap-

proach. First, the evaluation of the detailed temporal modelling of our audio features shows that, when extracted from polyphony, the resulting descriptors are not very discriminative between different instruments. We could not observe any improvement in performance when they were incorporated for the drum recognition task, even if a majority of the selected descriptors are describing fine temporal characteristics (see Fig. 3). Moreover, hardly any of these descriptors are selected for the pitched model. This implies that in the context of polyphony a detailed modelling becomes impossible for both short (percussive) and longer time-scale analysis (pitched), and that the remaining temporal aspects are best encoded in the coarse delta descriptors. That all strengthens the fact that temporal information is important for recognition but also shows the problems of modelling it accurately. These outcomes partly conform with the results presented in [11] by identifying fine temporal modelling of polyphonic audio as very fragile descriptors but proving the more coarse delta coefficients to be powerful, even when extracted from polyphonic data. Secondly, even if there would be some headroom for improvements, the algorithm will never be able to solve certain, dead-easy for humans, recognition tasks. Therefore we clearly see the need for different approaches in this area, starting from new audio representations to new algorithms for polyphonic processing. Enhanced signal processing as a front end system coupled with a complete probabilistic architecture (both bottom-up and top-down) could help to discover new paths, where an explicit source separation is not needed. Moreover, the integration of different knowledge sources could increase performance, as one solution might not always be applicable to all problems at hand.

6. CONCLUSIONS

In this paper we addressed three open gaps in automatic recognition of instruments from polyphonic audio. First we showed that by providing extensive, well designed datasets, statistical models are *scalable* to commercially available polyphonic music. Second, to account for *instrument generality*, we presented a consistent methodology for the recognition of 11 pitched and 3 percussive instruments in the main western genres classical, jazz and pop/rock. Finally, we examined the importance and modelling accuracy of *temporal characteristics* in combination with statistical models. Thereby we showed that modelling the temporal behaviour of raw audio features improves recognition performance, even though a detailed modelling is not possible. Results showed an average classification accuracy of 63% and 78% for the pitched and percussive recognition task, respectively. Although no complete system was presented, the developed algorithms could be easily incorporated into a robust recognition tool, able to index unseen data or label query songs according to the instrumentation.

ACKNOWLEDGEMENTS

The authors want to thank Joan Serrà and Emilia Gómez for their help in improving the quality and style of this publication. This research has been partially funded by the EU-IP project PHAROS [1] , IST-2006-045035.

7. REFERENCES

[1] X. Serra, R. Bresin, and A. Camurri, "Sound and music computing: Challenges and strategies," *Journal of New Music Research*, vol. 36, no. 3, pp. 185–190, 2007.

[2] S. McAdams, S. Winsberg, S. Donnadieu, G. DeSoete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes," *Psychological Research*, vol. 58, no. 3, pp. 177–192, 1995.

[3] D. FitzGerald and J. Paulus, "Unpitched percussion transcription," in *Signal Processing Methods for Music Transcription*, pp. 131–162, Springer, 2006.

[4] P. Herrera, A. Klapuri, and M. Davy, "Automatic classification of pitched musical instrument sounds," in *Signal Processing Methods for Music Transcription*, pp. 163–200, Springer, 2006.

[5] T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. Okuno, "Instrument identification in polyphonic music: Feature weighting to minimize influence of sound overlaps," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–16, 2007.

[6] M. Every, "Discriminating between pitched sources in music audio," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 267–277, 2008.

[7] S. Essid, G. Richard, and B. David, "Instrument recognition in polyphonic music based on automatic taxonomies," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 68–80, 2006.

[8] J. Paulus and A. Klapuri, "Combining temporal and spectral features in hmm-based drum transcription," in *Proc. of ISMIR*, 2007.

[9] O. Gillet and G. Richard, "Transcription and separation of drum signals from polyphonic music," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 529–540, 2008.

[10] D. Little and B. Pardo, "Learning musical instruments from mixtures of audio with weak labels," in *Proc. of ISMIR*, 2008.

[11] J. Aucouturier and F. Pachet, "The influence of polyphony on the dynamical modelling of musical timbre," *Pattern Recognition Letters*, pp. 654–661, 2007.

[12] O. Gillet and G. Richard, "ENST-drums: an extensive audio-visual database for drum," in *Proc. of ISMIR*, 2006.

[13] K. Tanghe, M. Lesaffre, S. Degroeve, M. Leman, B. D. Baets, and J. Martens, "Collecting ground truth annotations for drum detection in polyphonic music," in *Proc. of ISMIR*, 2005.

[14] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," tech. rep., IRCAM, 2004.

[15] M. Haro, "Detecting and describing percussive events in polyphonic music," Master's thesis, Universitat Pompeu Fabra, Spain, 2008.

[16] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proc. of Int. Conf. on Machine Learning*, pp. 359–366, 2000.

[17] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," *Annals of Statistics*, pp. 451–471, 1998.

[18] P. H. Kvam and B. Vidakovic, *Nonparametric Statistics with Applications to Science and Engineering*. Wiley, 2007.

[1] `http://www.pharos-audiovisual-search.eu`