Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques

Perfecto Herrera, Alexandre Yeterian, Fabien Gouyon

Universitat Pompeu Fabra
Pg. Circumval.lació 8
08003 Barcelona, Spain
+34 935422806
{pherrera, ayeter, fgouyon}@iua.upf.es

Abstract. We present a comparative evaluation of automatic classification of a sound database containing more than six hundred drum sounds (kick, snare, hihat, toms and cymbals). A preliminary set of fifty descriptors has been refined with the help of different techniques and some final reduced sets including around twenty features have been selected as the most relevant. We have then tested different classification techniques (instance-based, statistical-based, and tree-based) using ten-fold cross-validation. Three levels of taxonomic classification have been tested: membranes versus plates (super-category level), kick vs. snare vs. hihat vs. toms vs. cymbals (basic level), and some basic classes (kick and snare) plus some sub-classes –i.e. ride, crash, open-hihat, closed hihat, high-tom, medium-tom, low-tom- (sub-category level). Very high hit-rates have been achieved (99%, 97%, and 90% respectively) with several of the tested techniques.

1. INTRODUCTION

Classification is one of the processes involved in audio content description. Audio signals can be classified according to miscellaneous criteria. A broad partition into speech, music, sound effects (or noises), and their binary and ternary combinations is used for video soundtrack descriptions. Sound category classification schemes for this type of materials have been recently developed [1], and facilities for describing sound effects have even been provided in the MPEG-7 standard [2]. Usually, music streams are broadly classified according to genre, player, mood, or instrumentation. In this paper, we do not deal with describing which instruments appear in a musical mixture. Our interest is much more modest as we focus only in deriving models for discrimination between different classes of isolated percussive sounds and, more specifically in this paper, of acoustic "standard" drum kit sounds (i.e. not electronic, not Latin, not brushed, etc.). Automatic labelling of instrument sounds has some obvious applications for enhancing sampling and synthesis devices' operating systems in order to help sound designers to categorize (or suggesting names for) new patches and samples. Additionally, we assume that some outcomes of research on this subject will be used for the more ambitious task of describing the instrumentation in a musical recording, at least of the "rhythm loop" type.

Previous research in automatic classification of sound from music instruments has focused in instruments with definite pitch. Classification of string and wind instrument sounds has been attempted using different techniques and features yielding to varying degrees of success (see [3] for an exhaustive review). Classification of percussive instruments, on the other hand, has attracted little interest from researchers. In one of those above cited studies with pitched sounds, Kaminskyj [4] included three pitched percussive categories (glockenspiel, xylophone and marimba) and obtained good classification results (ranging from 75% to 100%) with a K-NN algorithm. Schloss [5] classified the stroke type of congas using relative energy from selected portions of the spectrum. He was able to differentiate between high-low sounds and open, muffled, slap and bass sounds. Using a K-means clustering algorithm, Bilmes [6] also was able to differentiate between sounds of three different congas. McDonald [7] used spectral centroid trajectories as classificatory features of sounds from percussive instruments. Sillanpää [8] used a representation of spectral shape for identification of the basic five categories of a drum kit: bass drum, snares, toms, hihats, and cymbals. His research was oriented towards transcription of rhythm tracks and therefore he additionally considered the case of identification of several simultaneous sounds. A database of 128 sounds was identified with 87% of accuracy for the case of isolated sounds. Performance dramatically dropped when there were two or three simultaneous sounds (respectively 49% and 8% for complete identification, though at least one of the sounds in the mixture was correctly identified all the times). In a subsequent study [9], the classification method used energy, Bark-frequency and log-time resolution spectrograms, and a fuzzy-c clustering of the original feature vectors into four clusters for each sound class. Weighted RMS-error fitting and an iterative spectral subtraction of models was used to match the test sounds against learnt models. Unfortunately, no systematic evaluation was presented this time. Goto and Murakoa [10] also studied drum sound classification in the context of source separation and beat tracking [11]. They implemented an “energy profile”-based snare-kick discriminator, though no effectiveness evaluation was provided. As a general criticism, in the previous research there is a lack of systematic evaluation of the different factors involved in automatic classification, and the databases are small to draw robust conclusions. A more recent study on this subject in the context of basic rhythmic pulse extraction [12] intended to be systematic, but also used a small database and a reduced set of descriptors.

Research on perceptual similarity of sounds is another area that provides useful information for addressing the problem of automatic classification of drum sounds. In perceptual studies, dis-similarity judgments between pairs of sounds are elicited from human subjects. With multidimensional scaling techniques, researchers find the dimensions that underlie to dis-similarity judgments. Even further, with proper comparison between those dimensions and physical features of sounds, it is possible to discover the links between perceptual and physical dimensions of sounds [13], [14], [15], [16]. A three dimensional perceptual space for percussive instruments (not including bells) has been hypothesized by Lakatos [17] (but also see [18]). This percussive perceptual space spans three related physical dimensions: log-attack time, spectral centroid and temporal centroid. Additional evidence supporting them has been gathered during the multimedia content description format standardization process (MPEG-7) and, consequently, they have been included in MPEG-7 as

descriptors for timbres [19]. Graphical interactive testing environments that are linked to specific synthesis techniques [20] seem to be a promising way for building higher-dimensional perceptual spaces.

From another area of studies, those focusing on characteristics of beaten objects, it seems that information about the way an object is hit is conveyed by the attack segment, whereas the decay or release segment conveys information about the shape and material of the beaten object [21]. Repp [22] found that different hand-clapping styles (palm-to-palm versus fingers-to-palm) correlated with different spectral envelope profiles. Freed [23] observed that the attack segment conveyed enough information for the subjects to evaluate the hardness of a mallet hit. Four features were identified as relevant for this information: energy, spectral slope, spectral centroid and the time-weighted average centroid of the spectrum. Kaltzky et al. [24] have got experimental results supporting the main importance of the decay part (specifically the decay rate) of a contact sound in order to identify the material of the beaten object.

In the next sections we will present the method and results of our study on automatic identification of drum sounds. First we will discuss the features we initially selected for the task and the ways for using the smallest set without compromising classification effectiveness. Some techniques consider relevance of descriptors without considering the classification algorithm in which they are being issued, but there are also attribute selection techniques that are linked to specific classification algorithms. We will compare both approaches with three different classification approaches: instance-based, statistical-based, and tree-based. Classification results for three taxonomic levels (super-category, basic level classes, and sub-categories) of drum-kit instruments will then be presented and discussed.

2. METHOD

2.1 Selection of sounds

A database containing 634 sounds was set up for doing this study. Distribution of sounds into categories is shown in Table 1. Sounds were drawn from different commercial sample CD's and CD-ROMs. The main selection criteria were that they belonged to acoustic drums with as little reverberation as possible, and without any other effect applied to them. Also different dynamics and different physical instruments were looked for. Specific playing techniques yielding dramatic timbral deviations from a "standard sound" such as brushed hits or rim-shots were discarded.

Table 1. Categories used and number of sounds (inside parentheses) included in each category

Super-category	Basic-level	Sub-category
Membranes (380)	Kick (115)	Kick (115)
	Snare (150)	Snare (150)
	Tom (115)	Low (42)
		Medium (44)
		High (29)
Plates (263)	Hihat (142)	Open (70)
		Closed (72)
	Cymbal (121)	Ride (46)
		Crash (75)

2.2 Descriptors

We considered descriptors or features belonging to different categories: attack-related descriptors, decay-related descriptors, relative energies for selected bands and, finally, Mel-Frequency Cepstral Coefficients and variances. An amplitude-based segmentator was implemented in order to get an estimation of the attack-decay boundary position, for then computing those descriptors that used this distinction. Analysis window size for the computation of descriptors was estimated after computation of Zero-Crossing Rate.

2.2.1 Attack-related descriptors

Attack Energy (1), Temporal Centroid (2), which is the temporal centre of gravity of the amplitude envelope, Log Attack-Time (3), which is the logarithm of the length of the attack, Attack Zero-Crossing Rate (4), and TC/EA (5), which is the ratio of the Temporal Centroid to the length of the attack.

2.2.2 Decay-related descriptors

Decay Spectral Flatness (6) is the ratio between the geometrical mean and the arithmetical mean (this gives an idea of the shape of the spectrum, if it's flat, the sound is more "white-noise"-like; if flatness is low, it will be more "musical"); Decay Spectral Centroid (7), which is the centre of gravity of the spectrum; Decay Strong Peak (8), intended to reveal whether the spectrum presents a very pronounced peak (the thinner and the higher the maximum of the spectrum is, the higher value takes this parameter); Decay Spectral Kurtosis (9), the 4^{th} order central moment (it gives clues about the shape of the spectrum: "peaky" spectra have larger kurtosis than scattered or outlier-prone spectra.), Decay Zero-Crossing Rate (10); "Strong Decay" (11), a feature built from the non-linear combination of the energy and temporal centroid of a frame (a frame containing a temporal centroid near its left boundary and strong energy is said to have a "strong decay"); Decay Spectral Centroid Variance (12); Decay Zero-Crossing Rate Variance (13); and Decay Skewness (14), the 3^{rd} order central moment (it gives indication about the shape of the spectrum in the sense that asymmetrical spectra tend to have large skewness values).

2.2.3 Relative energy descriptors

By dividing the spectrum of the decay part into 8 bands of frequency, the energy lying in them was calculated, and then the relative energy percent for each band was computed. These bands were basically chosen empirically, according to the observations of several spectra from relevant instruments. The boundaries were fixed after several trials in order to get significant results, and were the following: 40-70 Hz. (15), 70-110 Hz. (16), 130-145 Hz. (17), 160-190 Hz. (18), 300-400 Hz. (19), 5-7 KHz. (20), 7-10 KHz. (21), and 10-15 KHz. (22).

2.2.4 Mel-Frequency Cepstrum Coefficients

MFCC's have been usually used for speech processing applications, though they have shown usefulness in music applications too [25]. As they can be used as a compact representation of the spectral envelope, their variance was also recorded in order to keep some time-varying information. 13 MFCC's were computed over the whole signal, and their means and variances were used as descriptors. In order to interpret the selected sets of features in section 3, we will use the numeric ID's 23-35 for the MFCC means, and 36-48 for the MFCC variances.

2.3 Classification techniques

We have selected three different families of techniques to be compared[1]: instance-based algorithms, statistical modelling with linear functions, and decision tree building algorithms. The *K-Nearest Neighbors* (K-NN) technique is one of the most popular for instance-based learning and there are several papers on musical instrument sound classification using K-NN [26], [27], [28] [4]. As a novelty in this research context, we have also tested another instance-based algorithm called *K** (pronounced "K-star"), which classifies novel examples by retrieving the nearest stored example using an entropy measure instead of an Euclidean distance. Systematic evaluations of this technique using standard test datasets [29] showed a significant improvement of performance over the traditional K-NN algorithm.

Canonical discriminant analysis is a statistical modelling technique that classifies new examples after deriving a set of orthogonal linear functions that partition the observation space into regions with the class centroids separated as far as possible, but keeping the variance of the classes as low as possible. It can be considered like an ANOVA (or MANOVA) that instead of continuous to-be-predicted variables uses discrete (categorical) variables. After a successful discriminant function analysis, "important" variables can be detected. Discriminant analysis has been successfully used by [30] for classification of wind and string instruments.

C4.5 [31] is a decision tree technique that tries to focus on relevant features and ignores irrelevant ones for partitioning the original set of instances into subsets with a

[1] The discriminant analysis was run with SYSTAT (http://www.spssscience.com/SYSTAT/), and the rest of analyses with the WEKA environment (www.cs.waikato.ac.nz/~ml/).

strong majority of one of the classes. Decision trees, in general, have been pervasively used for different machine learning and classification tasks. Jensen and Arnspang [32] or Wieczorkowska [33] have used decision trees for musical instrument classification. An interesting variant of C4.5, that we have also tested, is PART (partial decision trees). It yields association rules between descriptors and classes by recursively selecting a class and finding a rule that "covers" as many instances as possible of it.

2.4 Cross-validation

For the forthcoming experiments the usual ten-fold procedure was followed: 10 subsets containing a 90% randomly selected sample of the sounds were selected for learning or building the models, and the remaining 10% was kept for testing them. Hit-rates presented below have been computed as the average value for the ten runs.

3. RESULTS

3.1 Selection of relevant descriptors

Two algorithm-independent methods for evaluating the relevance of the descriptors in the original set have been used: Correlation-based Feature Selection (hence CFS) and ReliefF. CFS evaluates subsets of attributes instead of evaluating individual attributes. A "merit" heuristic is computed for every possible subset, consisting of a ratio between how predictive a group of features is and how much redundancy or inter-correlation there is among those features [34]. Table 2 shows the CFS-selected features in the three different contexts of classification we are dealing with. Note that a reduction of more than fifty percent can be achieved in the most difficult case, and that the selected sets for basic level and for sub-category classification show an important overlap.

Table 2. Features selected by the CFS method

Super-category	[21, 4, 22]
Basic-level	[2, 4, 5, 6, 7, 9, 10, 14, 15, 16, 17, 18, 20, 21, 22, 26, 27, 30, 39]
Sub-category	[1, 2, 3, 4, 5, 6, 7, 9, 10, 14, 15, 16, 17, 18, 19, 20, 21, 22, 26, 30, 39]

ReliefF evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and for the nearest different class [34]. Table 3 shows the ReliefF-selected features in the three different contexts. Note that the list is a ranked one –from most to least relevant- and that we have matched the cardinality of this list to the one yielded by the previous method, in order to facilitate their comparisons.

Table 3. Features selected by the ReliefF method

Super-category	[9, 14, 7]
Basic-level	[9, 14, 7, 19, 10, 17, 4, 25, 18, 6, 15, 21, 20, 16, 24, 26, 30, 31, 13]
Sub-category	[9, 14, 19, 7, 17, 10, 4, 25, 16, 15, 18, 6, 21, 20, 24, 30, 26, 31, 13, 28, 2]

Comparing the two methods it can be seen that all selected subsets for basic-level or for sub-category share more than 60% of features (but surprisingly they do not coincide at all when the target is the super-category). It is also evident that they include quite a heterogeneous selection of descriptors (some MFCC's, some energy bands, some temporal descriptors, some spectral descriptors…).

Contrasting with the previous "filtering" approaches, we also tested a "wrapping" approach for feature selection [35]. This means that features are selected in connection with a given classification technique which acts as a wrapper for the selection. Canonical Discriminant Analysis provides numerical indexes in order to decide about the relevance of a feature (but after analysis, not prior to it) as for example the F-to-remove value, or the descriptor's coefficients inside the canonical functions. For feature selection inside CDA it is usual to follow a stepwise (usually backwards) procedure. This strategy, however, only grants a locally optimal solution, so that an exhaustive (but sometimes impractical) search of all the combinations is recommended [36]. In our case, we have proceeded with a combination of backward stepwise plus some heuristic search. Table 4 shows the selected subsets, with the features ranked according the F-to-remove value (the most relevant first). A difference related to the filtering approaches is that with CDA the selected sets are usually larger. A large proportion of the selected features, otherwise, match those selected with the other methods.

Table 4. Features selected after Canonical Discriminant Analyses

Super-category	[4, 13, 19, 20, 37, 39]
Basic-level	[15, 9, 4, 20, 14, 2, 13, 26, 27, 3, 19, 8, 21, 39, 6, 11, 38]
Sub-category	[16, 15, 3, 9, 2, 17, 20, 13, 14, 19, 27, 26, 39, 7, 12, 10, 8, 37, 38, 4, 21, 22, 25, 33, 30, 29, 5, 24, 28, 45, 36, 34]

3.1 Classification results

We tested the three algorithms using the different subsets discussed in the previous section. Three different levels of classification were tested: super-category (plates versus membranes), basic-level (the five instruments) and sub-category (kick and snare plus some variations of the other three instruments: open and closed hihat, low, mid and high tom, crash and ride cymbal). Tables 5, 6 and 7 summarize the main results regarding hit rates for the three different classification schemes we have tested. Rows contain the different algorithms and columns contain the results using the different sets of features that were presented in the previous section. For the C4.5, the

number of leaves appears inside parentheses. For PART, the number of rules appears inside parentheses. The best method for each feature set has been indicated with bold type and the best overall result appears with grey background.

Table 5. Super-category classification hit rates for the different techniques and feature selection methods

	All features	CFS	ReliefF	CDA
K-NN (k=1)	**99.2**	97.9	93.7	96.7
K*	98.6	97.8	**94.8**	96.7
C4.5	97.2 (8)	98.6 (8)	**94.8** (12)	95.1(14)
PART	98.4 (5)	98.2 (6)	94.4 (6)	95.1(9)
CDA	99.1	94.7	88.1	**99.3**

Table 6. Basic-level classification hit rates for the different techniques and feature selection methods

	All features	CFS	ReliefF	CDA
K-NN (k=1)	96.4	95	95.6	95.3
K*	**97**	**96.1**	**97.4**	**95.8**
C4.5	93 (20)	93.3(21)	92.2(23)	94.2(18)
PART	93.3 (12)	93.(11)	93.1(11)	93.6(12)
CDA	92	93	91	95.7

Table 7. Sub-category classification hit rates for the different techniques and feature selection methods

	All features	CFS	ReliefF	CDA
K-NN (k=1)	**89.9**	87.7	89.4	87.9
K*	**89.9**	**89.1**	**90.1**	**90.7**
C4.5	82.6 (40)	83 (38)	81 (45)	85(43)
PART	83.3 (24)	84.1(27)	81.9 (29)	84.3(27)
CDA	82.8	86	82	86.6

A clear interaction effect between feature selection strategy and algorithm family can be observed: for instance-based algorithms ReliefF provides the best results while for the decision-trees the best results have been obtained with CFS. In the case of decision trees, selecting features with CFS is good not only for improving hit-rates but also for getting more compact trees, (i.e. with a small number of leaves and therefore smaller in size). As expected, the CDA-selected features have yielded the best hit-rates for the CDA, but surprisingly they have also yielded the best hit-rates for most of the decision-trees.

It is interesting to compare the results obtained using feature selection with those obtained with the whole set of features. For the super-category classification it seems

that all the selection procedures have operated an excessive deletion and performance has degraded up to 4% when using a selected subset. Note however that in this classification test the best overall result (CDA features with CDA classification) outperforms any of the figures obtained with the whole subset. For the basic-level and sub-category tests, the reduction of features degrades the performance of instance-based methods (but less than 1%), whereas it improves the performance of the rest.

After comparing families of algorithms it is clear that differences between them increase as the task difficulty increases. It is also evident that the best performance is usually found in instance-based ones (and specifically K* yields slightly better results than a simple K-NN), whereas tree-based yield the worst figures and CDA lies in between. Although decision trees do not provide the best overall performance, they have an inherent advantage over instance-based: expressing relationships between features and classes in terms of conditional rules. Table 8 exemplifies the type of rules that we get after PART derivation.

Table 8. Some of the PART rules for classification at the "basic-level". Correctly and wrongly classified instances are shown inside parentheses. We have left out some less general rules for clarity

SKEWNESS > 4.619122 AND B40HZ70HZ > 7.784892 AND MFCC3 <= 1.213368: Kick (105.0/0.0)	
	SPECCENTROID > 11.491498 AND
KURTOSIS > 26.140138 AND	B1015KHZ > 0.791702: HH (100.0/2.0)
TEMPORALCE <= 0.361035 AND ATTZCR > 1.478743: Tom (103.0/0.0)	
	SKEWNESS <= 4.485531 AND B160HZ190HZ <= 5.446338 AND
B710KHZ <= 0.948147 AND	MFCC3VAR > 0.212043 AND
KURTOSIS <= 26.140138 AND ATTZCR <= 22.661397: Snare (133.0/0.0)	MFCC4 > -0.435871: Cymbal (110.0/3.0)

Regarding CDA, an examination of the canonical scores plots provides some graphical hints about the performance of the four canonical discriminant functions needed for the basic-level case: the first one separates toms+kicks from hihats+cymbals, the second one separates the snare from the rest, the third one separates cymbals from hihats, and the fourth one separates toms from kicks. It should be noted that in the other cases it is more difficult to assign them a clear role.

Inspecting the confusion matrix for the instrument test, most of the errors consist in confusing cymbals with hihat, and tom with kick (and their inverse confusions, though with a lesser incidence). For the sub-instrument test, 60% of the misclassifications appear to be intra-category (i.e. between crash and ride, between open and closed hihat, etc.), and they are evenly distributed.

4. DISCUSSION

We have achieved very high hit rates for the automatic classification of standard drum sounds into three different classification schemes. The fact that, in spite of using three

very different classification techniques, we have obtained quite similar results could mean that the task is quite an easy one. It is true that the number of categories we have used has been small even for the most complex classification scheme. But it should also be noted that there are some categories that, at least from a purely perceptual point of view, do not seem to be easily separated (for example, low-toms from some kicks, or some snares from mid-toms or from some crash cymbals). Therefore, a contrasting additional interpretation for this good performance is to consider that our initial selection of descriptors was good. This statement gets support by the fact that the all-feature results are not much worse than results after feature selection. In the case of having a bad initial set, those bad features would have contributed to worsen the performance. As it has not been the case, we can conclude that from a good set of initial features, some near-optimal sets have been identified with the help of filtering or wrapping techniques. Most of the best features found can be considered as spectral descriptors: skewness, kurtosis, centroid, MFCC's. We included a very limited number of temporal descriptors, but, as expected, apart from ZCR, they do not seem to be needed for precise instrument classification.

In the section of improvements for subsequent research we may list the following: (1) A more systematic approach to description in terms of energy bands (for example, using Bark measures); (2) Evaluation of whole-sound descriptors against attack-decay decomposed descriptors (i.e. the ZCR); (3) Non-linear scaling of some feature dimensions; (4) Justified deletion of some observations (after analyzing the models, it seems that some outliers that contribute to the increment of the confusion rates should be considered as "bad" examples for the model because of audio quality or wrong class adscription).

5. CONCLUSIONS

In this study, we have performed a systematic study of the classification of standard drum sounds. After careful selection of descriptors and its refinement with different techniques, we have achieved very high hit-rates in three different classification tasks: super-category, basic-level category, and sub-category. In general, the most relevant descriptors for them seem to be ZCR, kurtosis, skewness, centroid, relative energy in specific bands, and some low-order MFCC's. Performance measures classification techniques have not yielded dramatic differences between classification techniques and therefore selecting one or another is clearly an application-dependent issue. We believe, though, that relevant performance differences will arise when more classes are included in the test, as we have planned for a forthcoming study. Regarding classification of mixtures of sounds, even if it is not yet clear if the present results will be useful, we have gathered interesting and relevant data in order to characterize different classes of drum sounds.

6. ACKNOWLEDGMENTS

The research reported in this paper has been partially funded by the EU-IST project CUIDADO.

REFERENCES

[1] Zhang, T. and Jay Kuo, C.-C.: Classification and retrieval of sound effects in audiovisual data management. In Proceedings of 33rd Asilomar Conference on Signals, Systems, and Computers (1991)

[2] Casey, M.A.: MPEG-7 sound recognition tools. IEEE Transactions on Circuits and Systems for Video Technology, 11, (2001) 37-747

[3] Herrera, P., Amatriain, X., Batlle, E., Serra, X.: A critical review of automatic musical instrument classification. In Byrd, D, Downie, J.S., and Crawford, T (Eds.), Recent Research in Music Information Retrieval: Audio, MIDI, and Score Kluwer Academic Press, in preparation.

[4] Kaminskyj, I.: Multi-feature Musical Instrument Sound Classifier. In Proceedings of Australasian Computer Music Conference (2001)

[5] Schloss,W.A.: On the automatic transcription of percussive music -from acoustic signal to high-level analysis. STAN-M-27. Stanford, CA, CCRMA, Department of Music, Stanford University (1985)

[6] Bilmes, J.: Timing is the essence: Perceptual and computational techniques for representing, learning and reproducing expressive timing in percussive rhythm. MSc, Thesis. Massachussetts Institute of Technology, Media Laboratory. Cambridge, MA. (1993)

[7] McDonald, S. and Tsang, C.P.: Percussive sound identification using spectral centre trajectories. In Proceedings of 1997 Postgraduate Research Conference (1997)

[8] Sillanpää, J.: Drum stroke recognition. Tampere University of Technology. Tampere, Finland (2000)

[9] Sillanpää, J., Klapuri, A., Seppänen, J., and Virtanen, T.: Recognition of acoustic noise mixtures by combined bottom-up and top-down approach. In Proceedings of European Signal Processing Conference, EUSIPCO-2000 (2000)

[10] Goto, M., Muraoka, Y.: A sound source separation system for percussion instruments. Transactions of the Institute of Electronics, Information and Communication Engineers D-II, J77, 901-911 (1994)

[11] Goto, M. and Muraoka, Y.: A real-time beat tracking system for audio signals. In Proceedings of International Computer Music Conference, 171-174 (1995)

[12] Gouyon, F. and Herrera, P.: Exploration of techniques for automatic labeling of audio drum tracks' instruments. In Proceedings of MOSART: Workshop on Current Directions in Computer Music (2001)

[13] Miller, J.R., Carterette, E.C.: Perceptual space for musical structures. Journal of the Acoustical Society of America, 58, 711-720 (1975)

[14] Grey, J.M.: Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America, 61, 1270-1277 (1977)

[15] McAdams, S., Winsberg, S., de Soete, G., and Krimphoff, J.: Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychological Research, 58, 177-192 (1995)

[16] Toiviainen, P., Kaipainen, M., and Louhivuori, J.: Musical timbre: Similarity ratings correlate with computational feature space distances. Journal of New Music Research, 282-298 (1995)

[17] Lakatos, S.: A common perceptual space for harmonic and percussive timbres. Perception and Psychophysics, 62, 1426-1439 (2000)

[18] McAdams, S., Winsberg, S.: A meta-analysis of timbre space. I: Multidimensional scaling of group data with common dimensions, specificities, and latent subject classes (2002)
[19] Peeters, G., McAdams, S., and Herrera, P.: Instrument sound description in the context of MPEG-7. In Proceedings of Proceedings of the 2000 International Computer Music Conference (2000)
[20] Scavone, G., Lakatos, S., Cook, P., and Harbke, C.: Perceptual spaces for sound effects obtained with an interactive similarity rating program. In Proceedings of International Symposium on Musical Acoustics (2001)
[21] Laroche, J., Meillier, J.-L.: Multichannel excitation/filter modeling of percussive sounds with application to the piano. IEEE Transactions onSpeech and Audio Processing, 2, (1994) 329-344
[22] Repp, B.H.: The sound of two hands clapping: An exploratory study. Journal of the Acoustical Society of America, 81, (1993) 1100-1109
[23] Freed, A.: Auditory correlates of perceived mallet hardness for a set of recorded percussive events. Journal of the Acoustical Society of America, 87, (1990) 311-322
[24] Klatzky, R.L., Pai, D.K., and Krotkov, E.P.: Perception of material from contact sounds. Presence: Teleoperators and Virtual Environments, 9, (2000) 399-410
[25] Logan, B.: Mel Frequency Cepstral Coefficients for Music Modeling. In Proceedings of International Symposium on Music Information Retrieval, ISMIR-2000. Plymouth, MA, (2000)
[26] Martin, K.D. and Kim, Y.E.: Musical instrument identification: A pattern-recognition approach. In Proceedings of Proceedings of the 136th meeting of the Acoustical Society of America. (1998)
[27] Fujinaga, I. and MacMillan, K.: Realtime recognition of orchestral instruments. In Proceedings of the 2000 International Computer Music Conference, (2000) 141-143
[28] Eronen, A.: Comparison of features for musical instrument recognition. In Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01) (2001)
[29] Cleary, J.G. and Trigg, L.E.: K*: An instance-based learner using an entropic distance measure. In Proceedings of International Conference on Machine Learning, (1995) 108-114
[30] Agostini, G., Longari, M., and Pollastri, E.: Musical instrument timbres classification with spectral features. In Proceedings of IEEE Multimedia Signal Processing Conference (2001)
[31] Quinlan,J.R.: C4.5: Programs for machine learning. Morgan Kaufmann. San Mateo, CA, (1993)
[32] Jensen, K. and Arnspang, J.: Binary decision tree classification of musical sounds. In Proceedings of the 1999 International Computer Music Conference. (1999)
[33] Wieczorkowska, A.: Classification of musical instrument sounds using decision trees. In Proceedings of the 8th International Symposium on Sound Engineering and Mastering, ISSEM'99, (1999) 225-230
[34] Hall, M.A.: Correlation-based feature selection for discrete and numeric class machine learning. In Proceedings of Seventeenth International Conference on Machine Learning (2000)
[35] Blum, A., Langley, P.: Selection of relevant features and examples in machine learning. Artificial Intelligence, 97, (1997) 245-271
[36] Huberty,C.J.: Applied discriminant analysis. John Wiley. New York (1994)