Problems with Automatic Classification of Musical Sounds

Alicja A. Wieczorkowska[1], Jakub Wróblewski[1],
Dominik Ślęzak[2,1], and Piotr Synak[1]

[1] Polish-Japanese Institute of Information Technology
ul. Koszykowa 86, 02-008 Warsaw, Poland
[2] Department of Computer Science University of Regina
Regina, SK, S4S 0A2, Canada

Abstract. Convenient searching of multimedia databases requires well annotated data. Labeling sound data with information like pitch or timbre must be done through sound analysis. In this paper, we deal with the problem of automatic classification of musical instrument on the basis of its sound. Although there are algorithms for basic sound descriptors extraction, correct identification of instrument still poses a problem. We describe difficulties encountered when classifying woodwinds, brass, and strings of contemporary orchestra. We discuss most difficult cases and explain why these sounds cause problems. The conclusions are drawn and presented in brief summary closing the paper.

1 Introduction

With the increasing popularity of multimedia databases, a need arises to perform efficient searching of multimedia contents. For instance, the user can be interested in finding specific tune, played by the guitar. Such searching cannot be performed efficiently on raw sound or image data. Multimedia data should be first annotated with descriptors that facilitate such search. Standardization of multimedia content description is a scope of MPEG-7 standard [9,13]. However, algorithms of descriptors extraction or database searching are not within a scope of this standard, so they are still object of research. This is why we decided to investigate labeling of mono sounds with the names of musical instruments that play these sounds.

In this paper, we deal with problems that arise when automatic classification of musical instrument sounds is performed. It is far from perfect, and we especially focus on these sounds (and instruments) that are misclassified.

2 Classification of Musical Instruments

There exist numerous musical instruments all over the world. One can group them into classes according to various criteria. Widely used Sachs-Hornbostel system [8] classifies musical instruments into the categories, which can be seen in Table 1.

Table 1. Categories and subcategories of musical instruments

Category	Criteria for subclasses	Subclasses	Instruments
idiophones	material	struck together	castanets
	whether pitch is important	struck	gongs
	no. of idiophones in instrument	rubbed	saw
	no. of resonators	scraped	washboards
		stamped	floors
		shaken	rattles
		plucked	Jew's harp
membranophones	whether has 1 or 2 heads	drums: cylindrical,	
	if there are snares, sticky balls	conical, barrel,	
	how skin is fixed on drum	hourglass, long,	
	whether drum is tuned	goblet,	darabukke
	how it is tuned	kettle, footed,	
	how it is played	frame drum	tambourine
	position of drum when played	friction drum	
	body material	mirliton/kazoo	kazoo
chordophones	number of strings	zither	piano
	how they are played	lute plucked	guitar
	tuning	lute bowed	violin
	presence of frets	harp	harps
	presence of movable bridges	lyre	
		bow	
aerophones	kind of mouthpiece:	flutes: side-blown,	
	blow hole	end-blown,	
	whistle	nose, multiple	
	single reed	globular flute	ocarina
	double reed	panpipes	
	lip vibrated	whistle mouthpiece	recorder
		single reed	clarinet
		double reed	oboe
		air chamber	accordion
		lip vibrated	brass
		free aerophone	bullroarers
electrophones			keyboards

Membranophones and idiophones are together called percussion. Contemporary classification also adds electrophones to this set. The categories are further divided into subcategories [17]. As one can see, the variety of instruments complicates the process of classifications, especially in case of percussion, when classification depends on the shape of the instrument. The sound parameterization for the classification purposes is often based on harmonic properties of sound, so dealing with definite pitch (fundamental frequency) sounds is much more common and convenient. This is why we decided to limit ourselves to instruments of definite pitch.

In our paper, we deal with chordophones and aerophones only. The instruments we analize include bowed lutes (violin, viola, cello, and double bass), side-blown flute, single reed, double reed, and lip vibrated. All of them produce sounds of definite pitch. We use fundamental frequency of musical instrument sounds as a basis of sound parameterization, as well as the envelope of the waveform. Further parameters are described in the next section.

3 Sound Parameterization

In our research, we dealt with 667 singular sounds of instruments, recorded from MUMS CDs with 44.1kHz frequency and 16bit resolution [15]. MUMS library is commonly used in experiments with musical instrument sounds [4–6,10,14,20], so we can consider them to be a standard.

There exist many descriptors that can be applied for the instrument classification purposes. Recently elaborated MPEG-7 standard for Multimedia Content Description Interface provides for 17 audio descriptors [9], but many other sound features have been applied by the researchers so far, including various spectral and temporal features, autocorrelation, cepstral coefficients, wavelet-based descriptors, and so on [1,2,5–7,10,11,14,16,18–20]. The sound parameterization we applied starts with extraction of the following temporal, spectral, and envelope descriptors [18]:

Temporal descriptors:

- *Length*: Signal length
- *Attack*, *Steady* and *Decay*: Relative length of the attack (till reaching 75% of maximal amplitude), quasi-steady (after the end of attack, till the final fall under 75% of maximal amplitude) and decay time (the rest of the signal), respectively
- *Maximum*: Moment of reaching maximal amplitude

Spectral descriptors:

- $EvenHarm$ and $OddHarm$: contents of even and odd harmonics in spectrum
- $Brightness$ and $Irregularity$ [12]:

$$Br = \frac{\sum_{n=1}^{N} nA_n}{\sum_{n=1}^{N} A_n} \quad Ir = \log \sum_{n=2}^{N-1} |20 \log A_n - \frac{20 \log(A_{n+1} A_n A_{n-1})}{3}| \quad (1)$$

  where A_N is the amplitude of nth partial (harmonic) and N is number of available partials
- $Tristimulus1, 2, 3$ [16]:

$$Tr_1 = \frac{A_1^2}{\sum_{n=1}^{N} A_n^2} \quad Tr_2 = \frac{\sum_{n=2,3,4} A_n^2}{\sum_{n=1}^{N} A_n^2} \quad Tr_3 = \frac{\sum_{n=5}^{N} A_n^2}{\sum_{n=1}^{N} A_n^2} \quad (2)$$

- $Frequency$: Fundamental frequency

Envelope descriptors:

- $Val_{Amp}1, \ldots, Val_{Amp}7$: Average values of amplitudes within 7 intervals of equal width for a given sound
- $EnvFill$: Area under the curve of envelope, approximated by means of values $Val_{Amp}1, \ldots, Val_{Amp}7$
- $Cluster$: Number of the closest of 6 representative envelope curves (obtained via clustering) shown in Figure 1 [18].

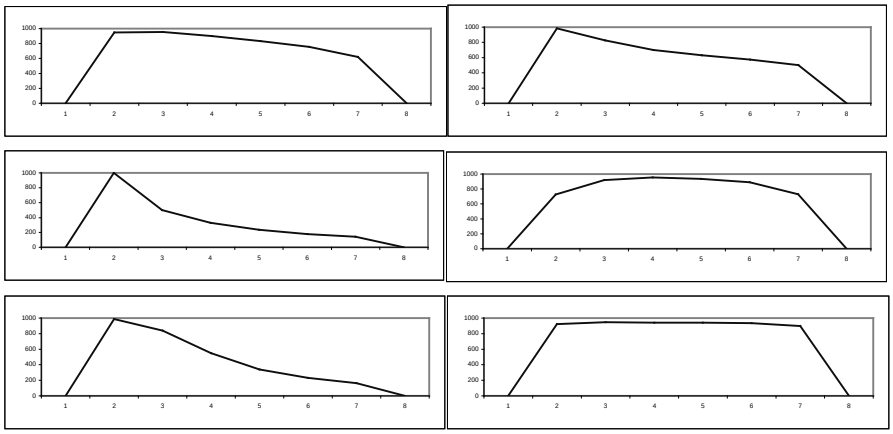

Fig. 1. The most typical shapes of sound envelopes, obtained as centroids of clusters

Some of the descriptors we calculated for the whole sound, whereas other where calculated for frames of length equal to 4 periods of sound and rectangular window through the whole sound (table WINDOW). The obtained basic data are represented as relational database. The structure of this database is shown in Figure 2 [18]. Objects of the database are classified according to both the instrument and articulation (how the sound is played) to the following classes: violin vibrato (denoted vln), violin pizzicato (vp), viola vibrato (vla), viola pizzicato (vap), cello vibrato (clv), cello pizzicato (clp), double bass vibrato (cbv), double bass pizzicato (cbp), flute (flt), oboe (obo), b-flat clarinet (cl), trumpet (tpt), trumpet muted (tpm), trombone (tbn), trombone muted (tbm), French horn (fhr), French horn muted (fhm) and tuba (tub).

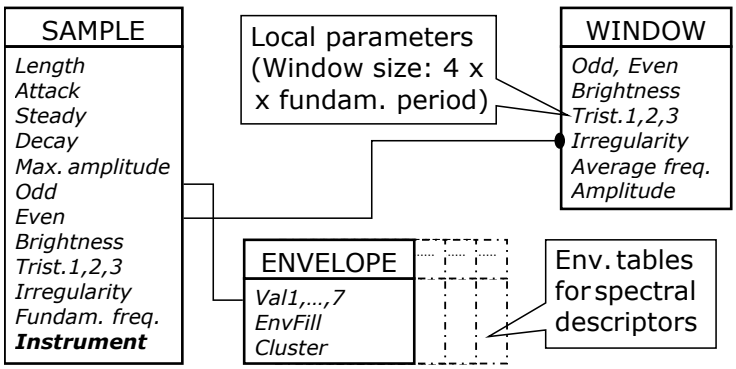

Fig. 2. Relational database of musical instrument sound descriptors

Apart from the data described above, we used new parameters, calculated for the basic ones. Namely, we extended the database by a number of new attributes defined as linear combinations of the existing ones. Additionally, the WINDOW table was used to search for the temporal templates, in order to use the found frequent episodes as new sound features [18].

4 Experiments

The data described in the previous section were used in experiments with automatic recognition of musical instrument sounds. The most common methods used in such experiments include k-nearest neighbor classifier (also with genetic algorithm to seek the optimal set of weights for the features), Bayes decision rules, decision trees, rough set based algorithms, neural networks, hidden Markov models and other classifiers [1,3,5,6,10,11,19,20]. Sometimes classification is performed in 2 stages:first, the sound is classified into a group (according to instrument category or articulation), and then the instrument is recognized. Extensive review of research in this domain is presented in [7].

Results of experiments vary, but apart from small data sets (for instance, 4 classes only), they are far from perfect, generally around 70-80% for instruments and about 90% for groups. The results are usually presented in form of correctness tables for various settings of the experiment method, and only some papers cover confusion matrices for the investigated instruments.

In the research for four woodwind instruments: oboe, sax, clarinet, and flute [3], confusions are presented in percentage for every pair. No overall pattern was observed for these data. In experiments for larger set of instruments (19), detailed confusion matrices are presented [10]. For these data, sax was frequently mistaken for clarinet (10 out of 37 samples for the best combined feature classifier) and trombone for French horn (7 out of 28 samples for the same classifier). Confusion matrix for our research is presented in Table 2.

Table 2. Confusion matrix for all (spectral and temporal) attributes

	cl	cbv	cbp	tpt	tpm	fhr	fhm	flt	obo	tbn	tbm	tub	vla	vap	clv	clp	vln	vp	%
cl	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
cbv	0	36	0	0	0	2	0	0	2	1	1	0	2	0	0	0	0	0	81.82
cbp	0	0	25	0	0	0	0	0	0	0	0	0	0	1	0	2	0	2	83.33
tpt	4	0	0	24	3	0	1	0	0	0	0	1	1	0	0	0	0	0	70.59
tpm	3	0	0	2	22	0	1	0	0	0	0	0	0	0	0	0	3	0	70.97
fhr	0	2	0	0	0	33	2	0	0	0	0	0	0	0	0	0	0	0	89.19
fhm	0	0	0	0	0	0	32	0	2	2	0	0	0	0	0	0	1	0	86.49
flt	3	0	0	0	0	0	0	33	0	1	0	0	0	0	0	0	0	0	89.19
obo	0	1	0	0	0	0	3	1	22	1	1	0	0	0	1	0	2	0	68.75
tbn	0	0	0	0	0	5	3	5	1	19	3	0	0	0	0	0	0	0	52.78
tbm	0	2	0	0	0	3	1	0	0	3	22	0	0	0	2	0	0	0	66.67
tub	0	1	0	0	0	0	0	1	0	0	0	29	0	0	1	0	0	0	90.63
vla	6	7	1	0	0	1	3	0	0	0	0	0	13	0	2	0	9	0	30.95
vap	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	4	0	9	61.76
clv	0	7	0	0	1	1	0	0	5	0	3	0	2	0	27	0	1	0	57.45
clp	0	0	6	0	0	0	0	0	0	1	0	0	0	13	0	18	0	1	46.15
vln	1	0	0	0	0	0	0	1	5	0	0	0	4	0	1	0	33	0	73.33
vp	0	0	1	0	0	0	0	0	0	0	0	0	0	14	0	0	0	25	62.50

As we can see, the most difficult instruments to classify in our case were violin and cello pizzicato, in 14 and 13 cases respectively, misclassified for viola

pizzicato. Since pizzicato sounds are very short and these instruments belong to the same family, this is not surprising. Clarinet was perfectly classified (we have not investigated sax that was problematic in [10]). Other instruments did not show such distinct patterns, but generally strings yielded lower results, with viola being the most difficult to classify correctly. Average recognition rate was 70.61%. These results are comparable with other research, and also with human achievements in musical instrument sound classification [3].

5 Conclusions

Classification of musical instrument sounds must take into account various articulation methods and categorization of instruments. In our research, we investigated sounds of non-percussion instruments of contemporary orchestra, including strings, woodwind, and brass. The most difficult were string sounds, especially when the investigated sounds are very short and change dramatically in time, i.e. played pizzicato (string plucked with finger). Such sounds are also difficult to parameterize, since analyzing frame must be also very short, and the sound features change very quickly. String sound played vibrato are also quite similar, so it is understandable that they can be mistaken. Generally, instruments belonging to the same category, or, even worse, to the same subcategory, are more difficult to discern. Additionally, vibration introduces fluent changes of sound features and also makes recognition more challenging. However, we hope that investigation of various sound parameterization techniques, combined with testing of various classification algorithms, may move forward the research on automatic indexing of musical sounds.

6 Acknowledgements

This research was partially supported by Polish National Committee for Scientific Research (KBN) in form of PJIIT Project No. $ST/MUL/01/2002$.

References

1. Batlle, E. and Cano, P. (2000) Automatic Segmentation for Music Classification using Competitive Hidden Markov Models. Proceedings of International Symposium on Music Information Retrieval. Plymouth, MA.
2. Brown, J. C. (1999) Computer identification of musical instruments using pattern recognition with cepstral coefficients as features. J. Acoust. Soc. of America, **105**, 1933–1941
3. Brown, J. C., Houix, O., and McAdams, S. (2001) Feature dependence in the automatic identification of musical woodwind instruments. J. Acoust. Soc. of America, **109**, 1064–1072
4. Cosi, P., De Poli, G., and Lauzzana, G. (1994) Auditory Modelling and Self-Organizing Neural Networks for Timbre Classification. Journal of New Music Research, **23**, 71–98

5. Eronen, A. and Klapuri, A. (2000) Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2000. Plymouth, MA. 753–756
6. Fujinaga, I. and McMillan, K. (2000) Realtime recognition of orchestral instruments. Proceedings of the International Computer Music Conference. 141–143
7. Herrera, P., Amatriain, X., Batlle, E., and Serra X. (2001) Towards instrument segmentation for music content description: a critical review of instrument classification techniques. In: Proc. of ISMIR 2000, Plymouth, MA
8. Hornbostel, Erich M. v. and Sachs, C. (1914) Systematik der Musikinstrumente. Ein Versuch. Zeitschrift für Ethnologie, **46**, (4-5):553–90. Available at http://www.uni-bamberg.de/ppp/ethnomusikologie/HS-Systematik/HS-Systematik
9. ISO/IEC JTC1/SC29/WG11 (2002) MPEG-7 Overview. Available at http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm
10. Kaminskyj, I. (2000) Multi-feature Musical Instrument Classifier. MikroPolyphonie **6** (online journal at http://farben.latrobe.edu.au/)
11. Kostek, B. and Czyzewski, A. (2001) Representing Musical Instrument Sounds for Their Automatic Classification. J. Audio Eng. Soc., **49(9)**, 768–785
12. Krimphoff, J., Mcadams, S., and Winsberg, S. (1994) Caractérisation du Timbre des Sons Complexes. II. Analyses acoustiques et quantification psychophysique. Journal de Physique IV, Colloque C5, J. de Physique III, 4, 3ème Congrès Français d'Acoustique, I, 625–628
13. Lindsay, A. T. and Herre, J. (2001) MPEG-7 and MPEG-7 Audio – An Overview. J. Audio Eng. Soc., **49(7/8)**, 589–594
14. Martin, K. D. and Kim, Y. E. (1998) 2pMU9. Musical instrument identification: A pattern-recognition approach. 136-th meeting of the Acoustical Soc. of America, Norfolk, VA
15. Opolko, F. and Wapnick, J. (1987) MUMS – McGill University Master Samples. CDs
16. Pollard, H. F. and Jansson, E. V. (1982) A Tristimulus Method for the Specification of Musical Timbre. Acustica, **51**, 162–171
17. SIL International (1999) 534 Musical Instruments subcategories. http://www.sil.org/LinguaLinks/Anthropology/ExpnddEthnmsclgyCtgrCltrlMtrls/MusicalInstrumentsSubcategorie.htm
18. Ślęzak, D., Synak, P., Wieczorkowska, A., and Wróblewski, J. (2002) KDD-based approach to musical instrument sound recognition. In Hacid M.-S., Raś Z., Zighed D. A., Kodratoff Y. (Eds.), Foundations of Intelligent Systems. Proc. 13th International Symposium ISMIS 2002. LNAI 2366, Springer, 29–37
19. Wieczorkowska, A. A. (1999) The recognition efficiency of musical instrument sounds depending on parameterization and type of a classifier (in Polish), Ph.D. Dissertation, Technical University of Gdańsk, Gdańsk.
20. Wieczorkowska, A. (1999) Rough Sets as a Tool for Audio Signal Classification. In Z. W. Ras, A. Skowron (Eds.), Foundations of Intelligent Systems, LNCS/LNAI 1609, Springer, 367–375