Parameter-Based Categorization for Musical Instrument Retrieval

Rory Lewis[1] and Alicja Wieczorkowska[2]

[1] University of North Carolina, 9201 University City Blvd. Charlotte, NC 28223, USA
[2] Polish-Japanese Institute of Information Technology, Koszykowa 86, 02-008 Warsaw, Poland

Abstract. In the continuing goal of codifying the classification of musical sounds and extracting rules for data mining, we present the following methodology of categorization, based on numerical parameters. The motivation for this paper is based upon the fallibility of Hornbostel and Sachs generic classification scheme, used in Music Information Retrieval for instruments. In eliminating the redundancy and discrepancies of Hornbostel and Sachs' classification of musical sounds we present a procedure that draws categorization from numerical attributes, describing both time domain and spectrum of sound. Rather than using classification based directly on Hornbostel and Sachs scheme, we rely on the empirical data describing the log attack, sustainability and harmonicity. We propose a categorization system based upon the empirical musical parameters and then incorporating the resultant structure for classification rules.

1 Instrument Classification

Information retrieval of musical instruments and their sounds has invoked a need to constructive cataloguing conventions with specialized vocabularies and other encoding schemes. For example the Library of Congress subject headings [1] and the German Schlagwortnormdatei Decimal Classification both use the Dewey classification system [3,11] In 1914 Hornbostel-Sachs devised a classification system, based on the Dewey decimal classification which essentially classified all instruments into strings, wind and percussion. Later it went further and broke instruments into four categories:

1.1 Idiophones, where sound is produced by vibration of the body of the instrument
2.2 Membranophones, where sound produced by the vibration of a membrane
3.3 Chordophones, where sound is produced by the vibration of strings
4.4 Aerophones, where sound is produced by vibrating air.

For purposes of music information retrieval, the Hornbostel-Sachs cataloguing convention is problematic, since it contains exceptions, i.e. instruments that could fall into a few categories. This convention is based on what element vibrates to produce sound (air, string, membrane, or elastic solid body), and playing method, shape, relationship of parts of the instrument and so on. Since

M. Kryszkiewicz et al. (Eds.): RSEISP 2007, LNAI 4585, pp. 784–792, 2007.

this classification follows a humanistic conventions, it makes it incompatible for a knowledge discovery discourse. For example, a piano emits sound when the hammer strikes strings. For many musicians, especially playing jazz, the piano is considered percussive, yet its the string that emits the sound vibrations, so it is classifies as a chordophone, according to Sachs and Hornbostel scheme. Also, the tamborine comprises a membrane and bells making it both an membranophone and an idiophone. Considering this, our paper presents a basis for an empirical music instrument classification system conducive for music information retrieval, specifically for automatic indexing of music instruments.

2 A Three-Level Empirical Tree

We focus on three properties of sound waves that can be calculated for any sound and can differentiate. They are: log-attack, harmonicity and sustainability. The first two properties are part of the set of descriptors for audio content description provided in the MPEG-7 standard and have aided us in musical instrument timbre description, audio signature and sound description [16]. The third one is based on observations of sound envelopes for singular sound of various instruments and for various playing method, i.e. articulation.

2.1 LogAttackTime (LAT)

The motivation for using the MPEG-7 temporal descriptor, LogAttackTime (LAT), is because segments containing short LAT periods cut generic percussive (and also sounds of plucked or hammered string) and harmonic (sustained) signals into two separate groups [6,7]. The *attack* of a sound is the first part of a sound, before a real note develops where the LAT is the logarithm of the time duration between the point where the signal starts to the point it reaches its stable part.[12] The range of the LAT is defined as $log_{10}(\frac{1}{samplingrate})$ and is determined by the length of the signal. Struck instruments, such a most percussive instruments have a short LAT whereas blown or vibrated instruments contain LATs of a longer duration.

$$LAT = log_{10}(T1 - T0), \tag{1}$$

where $T0$ is the time the signal starts; and $T1$ is reaches its sustained part (harmonic space) or maximum part (percussive space).

2.2 AudioHarmonicityType (HRM)

The motivation for using the MPEG-7 descriptor, AudioHarmonicityType is that it describes the degree of harmonicity of an audio signal.[7] Most "percussive" instruments contain a latent indefinite pitch that confuses and causes exceptions to parameters set forth in Hornbostel-Sachs. Furthermore, some percussive instruments such as a cuica or guido contain a weak LogAttackTime and therefore fall

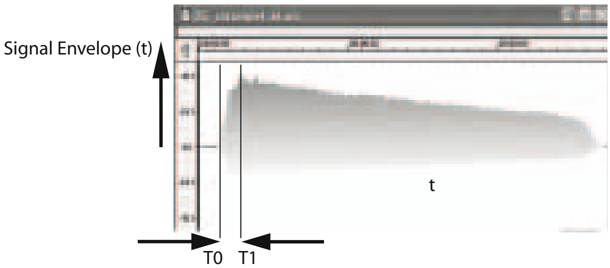

Fig. 1. Illustration of log-attack time. *T0* can be estimated as the time the signal envelope exceeds .02 of its maximum value. *T1* can be estimated, simply, as the time the signal envelope reaches its maximum value.

into non-percussive cluster while still maintaining an indefinite pitch (although, we can perceive differences in contents of low and high frequencies in percussive sounds as well). The use of the descriptor AudioHarmonicityType theoretically should solve this issue. It includes the weighted confidence measure, SeriesOfScalarType that handles portions of signal that lack clear periodicity. AudioHarmonicity combines the ratio of harmonic power to total power: HarmonicRatio, and the frequency of the inharmonic spectrum: UpperLimitOfHarmonicity.

First: We make the Harmonic Ratio *H(i)* the maximum *r(i,k)* in each frame, i where a definitive periodic signal for *H(i)* =1 and conversely white noise = 0.

$$H(i) = max\ r(i,k) \tag{2}$$

where *r(i,k)* is the normalised cross correlation of frame i with lag k:

$$r(i,k) = \sum_{j=m}^{m+n-1} s(j)\, s(j-k) \Bigg/ \left(\sum_{j=m}^{m+n-1} s(j)^2 * \sum_{j=m}^{m+n-1} s(j-k)^2 \right)^{\frac{1}{2}} \tag{3}$$

where s is the audio signal, m=i*n, where i=0, $M-1$=frame index and M = the number of frames, n=t*sr, where t = window size (10ms) and sr = sampling rate, k=*1*, K=*lag*, where K=ω*sr, ω = maximum fundamental period expected (40ms)

Second: Upon obtaining the i) DFTs of $s(j)$ and comb-filtered signals $c(j)$ in the AudioSpectrumEnvelope and ii) the power spectra $p(f)$ and $p'(f)$ in the AudioSpectrumCentroid we take the ratio f_{lim} and calculate the sum of power beyond the frequency for both $s(j)$ and $c(j)$:

$$a(f_{lim}) = \sum_{f=f_{lim}}^{f_{max}} p'(f) \Bigg/ \sum_{f=f_{lim}}^{f_{max}} p(f) \tag{4}$$

where f_{max} is the maximum frequency of the DFT.

Third: Starting where $f_{lim} = f_{max}$ we move down in frequency and stop where the greatest frequency, f_{ulim}'s ratio is smaller than 0.5 and convert it to an octave scale based on 1 kHz:

$$UpperLimitOfHarmonicity = log2(f_{ulim}/1000) \quad (5)$$

2.3 Sustainability (S)

We define sustainability into 5 categories based on the degree of dampening or sustainability the instrument can maintain over a maximum period of 7 seconds. For example, a flutist, horn player and violinist can maintain a singular note for more than 7 seconds therefore they receive a 1. Conversely a plucked guitar or single drum note typically cannot sustain that one sound for more than 7 seconds. It is true that a piano with pedal could maintain a sound after ten seconds but the sustainability factor would be present.

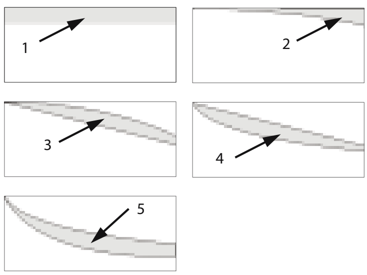

Fig. 2. Five levels of sustainability to severe dampening

3 Experiments

The sound data consists of a sample set of 156 signals extracted from our online database at http://www.mir.uncc.edu which contains 6,300 segmented sounds mostly from MUMS audio CD's that contain samples of broad range of musical instruments, including orchestral ones, piano, jazz instruments, organ, etc. [10] These CD's are widely used in musical instrument sound research [2,9,15,5,8,4], so they can be considered as a standard. The database consists of 188 samples each representing just one sample from group that make up the 6,300 files in the database. Mums divides the database into the following 18 classes: violin vibrato, violin pizzicato, viola vibrato, viola pizzicato, cello vibrato, cello pizzicato, double bass vibrato, double bass vibrato, double bass pizzicato, flute, oboe, b-flat clarinet, trumpet, trumpet muted, trombone, trombone muted, French horn, French horn muted, and tuba. Preprocessing these groups is not a part of rough set theory because rough sets require that input data process the rough sets. Rough set are objective with respect to its data. Here we discretize, using MPEG-7 classifiers as the experts. This is the point of the paper, we show a novel, empirical methodology of dividing sounds conducive to automatic retrieval of music.

4 Testing

The principle objective of our testing is to prove how parameter-based classification differs, and when used on Sachs-Hornbostel - improves Sachs-Hornbostel. Our parameters are machine-based, based on MPEG-7 and the temporal signal dampening. It is not based upon humanistic intuitiveness. We first prove that our attributes divide instruments into groups. Next we prove that our objects, which are in leaves for a given class, actually represent another class. This will show how parameter-based classification differs from and improves Sachs-Hornbostel. To induce the classification rules in the form of decision trees from a set of given examples we used Quinlan's C4.5 algorithm. [13] The algorithm constructs a decision tree to form production rules from an unpruned tree. Next a decision tree interpreter classifies items which produces the rules. We used Bratko's Orange software [14] and implement C4.5 with scripting in Python.

4.1 HRM, LAT, S, with HS01

The first test comprised the testing of the decision attribute Sachs-Hornbostel-level-1 against our two MPEG-7 descriptors, Harmonicity (HRM), Log Attack (LAT) and our temporal feature Sustainability (S). The Sachs-Hornbostel-level-1 attribute consists of four classes based upon human intuitiveness: aerophones, idiophones, chordophones and membranophones. See Appendix Figure 3

4.2 HRM, LAT, S, with HS02

The second test comprised the testing of the decision attribute Sachs-Hornbostel-level-2 against the HRM, LAT and S descriptors. The Sachs-Hornbostel-level-2 attribute consists of four classes: aerophones, idiophones, chordophones and membranophones. See Appendix Figure 4

4.3 HRM, LAT, S, with Instruments

The third test comprised the testing of the decision attribute instruments against the HRM, LAT and S descriptors. The Instrument attribute consists of four classes that describe instruments in the manner machines look at their signals: percussion, blown, string and struck Harmonics. See Appendix Figure 5

4.4 Resulting Tree

The resulting tree shows how the sound objects are grouped, and we can compare how this classification differs from Sachs-Hornbostel system. The misclassified objects show discrepancies between the Sachs-Hornbostel system, and sound properties described by physical attributes. The novelty of this methodology is that adding the temporal feature and grouping the instruments from the machines point of view have lead to 83% correctness. We have 26 more MPEG-7 descriptors to use with this methodology to breakdown the 17% misclassified

5 Summary and Conclusion

The idea and experiments presented in this paper show how musical instrument sounds can be classified according to physical properties of sounds, described by numerical parameters. The differences between obtained classification and Sachs-Hornbostel classification system show how ambiguous sounds, representing instruments played with various articulation, can be unambiguously classified.

We plan to continue our experiments, using more of our MPEG-7 features and applying clustering algorithms in order to find probably better classification scheme for musical instrument sounds.

References

1. Brenne, M.: Storage and retrieval of musical documents in a FRBR-based library catalogue: Thesis, Oslo University College Faculty of journalism, library and information science (2004)
2. Cosi, P., De Poli, G., Lauzzana, G.: Auditory Modelling and Self-Organizing Neural Networks for Timbre Classification. Journal of New Music Research 23, 71–98 (1994)
3. Doerr, M.: Semantic Problems of Thesaurus Mapping. Journal of Digital Information 1(8), Article No. 52, 2001-03-26, 2001–03 (2001)
4. Eronen, A., Klapuri, A.: Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2000, Plymouth, MA, pp. 753–756 (2000)
5. Fujinaga, I., McMillan, K.: Realtime recognition of orchestral instruments. In: Proceedings of the International Computer Music Conference, pp. 141–143 (2000)
6. Gomez, E., Gouyon, F., Herrera, P., Amatriain, X.: Using and enhancing the current MPEG-7 standard for a music content processing tool. In: Proceedings of the 114th Audio Engineering Society Convention, March 2003, The Netherlands, Amsterdam (2003)
7. Information Technology — Multimedia Content Description Interface — Part 4: Audio. ISO/IEC JTC 1/SC 29, Date: 2001-06-9. ISO/IEC FDIS 15938-4:2001(E) ISO/IEC J/TC 1/SC 29/WG 11 Secretariat: ANSI (2001)
8. Kaminskyj, I.: Multi-feature Musical Instrument Classifier. MikroPolyphonie, 6 (2000) (online journal at `http://farben.latrobe.edu.au/`)
9. Martin, K.D., Kim, Y.E.: 2pMU9. Musical instrument identification: A pattern-recognition approach. 136-th meeting of the Acoustical Soc. of America, Norfolk, VA (1998)
10. Opolko, F., Wapnick, J.: MUMS – McGill University Master Samples. CD's (1987)
11. Patel, M., Koch, T., Doerr, M., Tsinaraki, C.: Semantic Interoperability in Digital Library Systems. IST-2002-2.3.1.12 Technology-enhanced Learning and Access to Cultural Heritage. UKOLN, University of Bath (2005)
12. Peeters, G., McAdams, S., Herrera, P.: Instrument sound description in the context of MPEG-7: In: Proceedings of the International Computer Music Conference (ICMC'00), Berlin, Germany (2000)
13. Quinlan, J.R.: 2pMU9. Bagging, boosting, and C4. 5. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence, vol. 725, p. 730 (1996)

14. Demsar, J., Zupan, B., Leban, G.: `http://www.ailab.si/orange`
15. Wieczorkowska, A.: Rough Sets as a Tool for Audio Signal Classification. In: Raś, Z.W., Skowron, A. (eds.) Foundations of Intelligent Systems. LNCS, vol. 1609, pp. 367–375. Springer, Heidelberg (1999)
16. Wieczorkowska, A., Wróblewski, J., Synak, P., Ślęzak, D.: Application of temporal descriptors to musical instrument sound recognition. In: Proceedings of the International Computer Music Conference (ICMC'04), Berlin, Germany (2004)

Appendix

Classification Tree	Class	P(Class)	P(Target)	#Inst	Rel. distr.	Abs. distr.
<root>	idio	42	19	156	19:21:42:19	0:0:0:0
S <2.000	aero	53	53	49	53:27:14:6	1:0:0:0
HAR <820899.438	mem	43	29	7	29:14:14:43	0:0:0:0
LAT <77207.203	mem	75	0	4	0:0:25:75	0:0:0:1
LAT >=77207.203	aero	67	67	3	67:33:0:0	1:0:0:0
HAR >=820899.438	aero	57	57	42	57:29:14:0	1:0:0:0
S <1.000	aero	72	72	25	72:24:4:0	1:0:0:0
HAR <989678.688	aero	56	56	16	56:38:6:0	1:0:0:0
LAT <-218904.000	aero	100	100	5	100:0:0:0	1:0:0:0
LAT >=-218904.000	chrd	55	36	11	36:55:9:0	0:1:0:0
LAT <22621.900	chrd	100	0	4	0:100:0:0	0:1:0:0
LAT >=22621.900	aero	57	57	7	57:29:14:0	1:0:0:0
HAR >=989678.688	aero	100	100	9	100:0:0:0	1:0:0:0
S >=1.000	aero	35	35	17	35:35:29:0	0:0:0:0
LAT <-343387.000	idio	56	0	9	0:44:56:0	0:0:1:0
HAR <982953.000	chrd	67	0	6	0:67:33:0	0:1:0:0
HAR >=982953.000	idio	100	0	3	0:0:100:0	0:0:1:0
LAT >=-343387.000	aero	75	75	8	75:25:0:0	1:0:0:0
S >=2.000	idio	54	4	107	4:18:54:24	0:0:1:0
LAT <-1182790.000	mem	52	9	44	9:7:32:52	0:0:0:1
HAR <938294.188	mem	55	11	38	11:0:34:55	0:0:0:1
S <4.000	mem	62	10	29	10:0:28:62	0:0:0:1
LAT <-1755700.000	idio	55	18	11	18:0:55:27	0:0:1:0
HAR <594878.750	idio	63	0	8	0:0:63:38	0:0:1:0
HAR >=594878.750	aero	67	67	3	67:0:33:0	1:0:0:0
LAT >=-1755700.000	mem	83	6	18	6:0:11:83	0:0:0:1
S >=4.000	idio	56	11	9	11:0:56:33	0:0:1:0
HAR <383054.438	mem	67	33	3	33:0:0:67	0:0:0:1
HAR >=383054.438	idio	83	0	6	0:0:83:17	0:0:1:0
HAR >=938294.188	chrd	50	0	6	0:50:17:33	0:1:0:0
LAT >=-1182790.000	idio	70	0	63	0:25:70:5	0:0:1:0
HAR <772931.313	idio	91	0	34	0:3:91:6	0:0:1:0
HAR >=772931.313	chrd	52	0	29	0:52:45:3	0:1:0:0
LAT <-485895.000	idio	62	0	21	0:33:62:5	0:0:1:0
LAT >=-485895.000	chrd	100	0	8	0:100:0:0	0:1:0:0

Fig. 3. C4.5 results testing the decision attribute Sachs-Hornbostel-level-1 against our two MPEG-7 descriptors, Harmonicity (HRM), Log Attack (LAT) and our temporal feature Sustainability (S). S is divided at the ¡2.000 and ¿=2.000 node, Harmonicity is divided at ¡820889 and ¿=820889 for S ¡2.000 whereas, at ¿=2.000 LAT cuts the tree at LAT ¡-1182790 and ¿=1182790.

Classification Tree	Class	P(Class)	P(Target)	#Inst	Rel. distr.
LAT >=77207.203	idio_struck	38	13	8	13:0:38:25:0:0:0:25:0:0:0:0:0:0:0:0:0
LAT <109305.000	mem_conical	67	0	3	0:0:0:67:0:0:0:33:0:0:0:0:0:0:0:0:0
LAT >=109305.000	idio_struck	60	20	5	20:0:60:0:0:0:0:20:0:0:0:0:0:0:0:0:0
HAR >=996751.688	idio_concussion	57	0	7	0:57:29:14:0:0:0:0:0:0:0:0:0:0:0:0:0
HAR <997782.063	idio_concussion	100	0	3	0:100:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0
HAR >=997782.063	idio_struck	50	0	4	0:25:50:25:0:0:0:0:0:0:0:0:0:0:0:0:0
S >=1.000	idio_struck	32	16	19	16:11:32:0:11:0:0:11:5:5:0:0:5:0:0:0:5
LAT <-313254.000	mem_cylindrical	27	27	11	27:9:9:0:18:0:0:18:0:9:0:0:9:0:0:0:0
LAT <-450396.000	mem_friction	25	13	8	13:13:13:0:25:0:0:25:0:13:0:0:0:0:0:0:0
LAT <-696760.000	mem_cylindrical	25	25	4	25:0:25:0:25:0:0:0:0:25:0:0:0:0:0:0:0
LAT >=-696760.000	chrd_composite	50	0	4	0:25:0:0:25:0:0:50:0:0:0:0:0:0:0:0:0
LAT >=-450396.000	mem_cylindrical	67	67	3	67:0:0:0:0:0:0:0:0:0:0:0:33:0:0:0:0
LAT >=-313254.000	idio_struck	63	0	8	0:13:63:0:0:0:0:0:13:0:0:0:0:0:0:0:13
S >=2.000	chrd_composite	22	7	107	7:5:21:2:0:1:1:22:3:3:8:3:2:7:11:3:3:0
HAR <605413.188	chrd_composite	33	7	46	7:2:4:4:0:0:0:33:0:0:11:2:4:9:17:4:2:0
LAT <-793303.000	chrd_composite	35	0	26	0:4:0:0:0:0:0:35:0:0:12:4:4:0:31:8:4:0
HAR <383054.438	chrd_composite	47	0	17	0:6:0:0:0:0:0:47:0:0:0:0:0:0:35:12:0:0
HAR >=383054.438	idio_shaken	33	0	9	0:0:0:0:0:0:0:11:0:0:33:11:11:0:22:0:11:0
S <4.000	aero_lip-vibrated	33	0	6	0:0:0:0:0:0:0:0:0:0:17:17:17:0:33:0:17:0
LAT <-1446330.000	aero_lip-vibrated	67	0	3	0:0:0:0:0:0:0:0:0:0:0:0:33:0:0:67:0:0:0
LAT >=-1446330.000	idio_shaken	33	0	3	0:0:0:0:0:0:0:0:0:0:33:0:33:0:0:0:33:0
S >=4.000	idio_shaken	67	0	3	0:0:0:0:0:0:0:33:0:0:67:0:0:0:0:0:0:0
LAT >=-793303.000	chrd_composite	30	15	20	15:0:10:10:0:0:0:30:0:0:10:0:5:20:0:0:0:0
LAT <-160552.000	mem_cylindrical	30	30	10	30:0:10:20:0:0:0:10:0:0:20:0:0:10:0:0:0:0
S <3.000	mem_cylindrical	33	33	3	33:0:0:0:0:0:0:0:0:0:33:0:0:33:0:0:0:0
S >=3.000	mem_cylindrical	29	29	7	29:0:14:29:0:0:0:14:0:0:14:0:0:0:0:0:0:0
HAR <268127.438	mem_conical	67	0	3	0:0:0:67:0:0:0:33:0:0:0:0:0:0:0:0:0
HAR >=268127.438	mem_cylindrical	50	50	4	50:0:25:0:0:0:0:0:0:0:25:0:0:0:0:0:0:0
LAT >=-160552.000	chrd_composite	50	0	10	0:0:10:0:0:0:0:50:0:0:0:0:10:30:0:0:0:0
HAR <536840.000	chrd_composite	67	0	6	0:0:17:0:0:0:0:67:0:0:0:0:0:17:0:0:0:0
HAR >=536840.000	aero_single-reed	50	0	4	0:0:0:0:0:0:0:25:0:0:0:0:25:50:0:0:0:0
HAR >=605413.188	idio_struck	33	7	61	7:7:33:0:0:2:2:15:5:5:7:3:0:5:7:2:3:0
S <4.000	idio_struck	37	6	52	6:8:37:0:0:2:2:15:6:4:4:0:0:6:6:2:4:0
HAR <856620.188	idio_struck	44	0	25	0:12:44:0:0:0:0:8:4:4:4:0:0:12:12:0:0:0
S <3.000	idio_struck	25	0	8	0:13:25:0:0:0:0:25:13:0:0:0:0:25:0:0:0:0
LAT <-960761.000	idio_struck	50	0	4	0:25:50:0:0:0:0:0:0:0:0:0:0:25:0:0:0:0
LAT >=-960761.000	chrd_composite	50	0	4	0:0:0:0:0:0:0:50:25:0:0:0:0:25:0:0:0:0
S >=3.000	idio_struck	53	0	17	0:12:53:0:0:0:0:0:0:6:6:0:0:6:18:0:0:0
HAR <807241.813	idio_struck	64	0	11	0:18:64:0:0:0:0:0:0:0:9:0:0:9:0:0:0:0
LAT <-696760.000	idio_struck	88	0	8	0:13:88:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0
LAT >=-696760....	idio_concussion	33	0	3	0:33:0:0:0:0:0:0:0:0:33:0:0:33:0:0:0:0
HAR >=807241.813	aero_lip-vibrated	50	0	6	0:0:33:0:0:0:0:0:0:17:0:0:0:0:50:0:0:0
HAR <833891.875	aero_lip-vibrated	67	0	3	0:0:0:0:0:0:0:0:0:33:0:0:0:0:67:0:0:0
HAR >=833891....	idio_struck	67	0	3	0:0:67:0:0:0:0:0:0:0:0:0:0:0:33:0:0:0
HAR >=856620.188	idio_struck	30	11	27	11:4:30:0:0:4:4:22:7:4:4:0:0:0:0:4:7:0
HAR <938294.188	chrd_composite	38	23	13	23:8:8:0:0:0:0:38:8:0:0:0:0:0:0:8:8:0
LAT <-1008950.000	mem_cylindrical	50	50	6	50:0:0:0:0:0:0:17:17:0:0:0:0:0:0:0:17:0
LAT >=-1008950.000	chrd_composite	57	0	7	0:14:14:0:0:0:0:57:0:0:0:0:0:0:0:14:0:0
HAR >=938294.188	idio_struck	50	0	14	0:0:50:0:0:7:7:7:7:7:7:0:0:0:0:0:7:0
S >=4.000	idio_shaken	22	11	9	11:0:11:0:0:0:0:11:0:11:22:22:0:0:11:0:0:0
LAT <-1153220.000	mem_frame	50	0	4	0:0:0:0:0:0:0:0:0:0:25:50:0:0:25:0:0:0
LAT >=-1153220.000	mem_cylindrical	20	20	5	20:0:20:0:0:0:0:20:0:20:20:0:0:0:0:0:0:0

Fig. 4. C4.5 results testing the decision attribute Sachs-Hornbostel-level-2 against our two MPEG-7 descriptors, Harmonicity (HRM), Log Attack (LAT) and our temporal feature Sustainability (S)

Classification Tree	Class	P(Class)	P(Target)	#Inst	Rel. distr.	Abs. distr.
<root>	percussion	58	58	156	58:17:17:8	1:0:0:0
S <2.000	blown	51	16	49	16:51:24:8	0:1:0:0
S <1.000	blown	63	10	30	10:63:23:3	0:1:0:0
HAR <506156.188	percussion	67	67	3	67:33:0:0	1:0:0:0
HAR >=506156.188	blown	67	4	27	4:67:26:4	0:1:0:0
HAR <989678.688	blown	56	0	18	0:56:39:6	0:1:0:0
LAT <-218904.000	blown	100	0	5	0:100:0:0	0:1:0:0
LAT >=-218904.000	string	54	0	13	0:38:54:8	0:0:1:0
LAT <22621.900	string	100	0	4	0:0:100:0	0:0:1:0
LAT >=22621.900	blown	56	0	9	0:56:33:11	0:1:0:0
HAR <942690.000	string	50	0	4	0:25:50:25	0:0:1:0
HAR >=942690....	blown	80	0	5	0:80:20:0	0:1:0:0
HAR >=989678.688	blown	89	11	9	11:89:0:0	0:1:0:0
S >=1.000	blown	32	26	19	26:32:26:16	0:0:0:0
LAT <-343387.000	percussion	33	33	9	33:0:33:33	0:0:0:0
LAT <-696760.000	percussion	50	50	4	50:0:50:0	1:0:1:0
LAT >=-696760.000	struck_Hrm	60	20	5	20:0:20:60	0:0:0:1
LAT >=-343387.000	blown	60	20	10	20:60:20:0	0:1:0:0
HAR <856620.188	percussion	67	67	3	67:33:0:0	1:0:0:0
HAR >=856620.188	blown	71	0	7	0:71:29:0	0:1:0:0
S >=2.000	percussion	78	78	107	78:1:14:7	1:0:0:0
HAR <772931.313	percussion	100	100	61	100:0:0:0	1:0:0:0
HAR >=772931.313	percussion	48	48	46	48:2:33:17	0:0:0:0
LAT <-485895.000	percussion	58	58	38	58:3:18:21	1:0:0:0
LAT <-1226300.000	percussion	87	87	15	87:7:7:0	1:0:0:0
LAT >=-1226300.000	percussion	39	39	23	39:0:26:35	0:0:0:0
S <4.000	struck_Hrm	40	30	20	30:0:30:40	0:0:0:0
S <3.000	percussion	67	67	3	67:0:0:33	1:0:0:0
S >=3.000	struck_Hrm	41	24	17	24:0:35:41	0:0:0:0
LAT <-671080.000	string	50	33	12	33:0:50:17	0:0:1:0
LAT <-1008...	string	60	0	5	0:0:60:40	0:0:1:0
LAT >=-100...	percussion	57	57	7	57:0:43:0	1:0:0:0
LAT >=-671080....	struck_Hrm	100	0	5	0:0:0:100	0:0:0:1
S >=4.000	percussion	100	100	3	100:0:0:0	1:0:0:0
LAT >=-485895.000	string	100	0	8	0:0:100:0	0:0:1:0

Fig. 5. C4.5 results testing of the decision attribute instruments against the HRM, LAT and S descriptors. The Class files indicate whether the instruments are percussive, blown, string or struck harmonics.