A STATISTICAL ANALYSIS OF TIMBRE DESCRIPTORS ON MUSICAL INSTRUMENT CLASSIFICATION

MUMT 621 Music Information Retrieval

12 April 2009

Song Hui Chon

ABSTRACT

Musical instrument classification is one of the biggest and the most popular problems in music information retrieval (MIR). The usual goal is to accurately classify sounds from one or more instruments according to the instrument family (such as string, wind or percussion). The classification is done using various descriptors (or features) that are computed from the sound stimuli based on information on pitch, timbre, excitation method and others. The zero-crossing rate, spectral centroid and mel-frequency cepstrum coefficients (MFCCs) are only three examples. A good number of previous works have concentrated on improving the classification performance by using a smart machine-learning algorithm on a set of previously proven efficient features, which are not necessarily related to timbre. This paper considers only the descriptors that are related to timbre perception, since timbre is one of the primary perceptive cues in human classification of musical instruments. Statistical analysis techniques were used for this approach to find out the best and the worst descriptors in two problems in musical instrument classification – instrument family classification and excitation classification in the wind instrument family. The result shows that there may be certain descriptors that are better and more reliable than others, with a moderately successful classification rate. The performance is expected to improve with an efficient combination of two or more "better" descriptors.

1. INTRODUCTION

Classification of musical instrument sounds is an important problem in MIR [Martin & Kim 1998] [Herrera-Boyer et al. 2003] [Livshin & Rodet 2004] [Wieczorkowska & Kolczynska 2008]. It is related to other major problems such as (multiple) fundamental estimation, timbre recognition & classification, solo instrument identification and automatic transcription. The main goal is maximization of the correct classification (and usually labeling) of the instrument sounds within the given criterion (e.g., the instrument family classification, the instrument recognition, or the excitation type classification) using a set of descriptors whose values are extracted from the input stimuli using the information on pitch, timbre and other aspects.

Timbre is a multidimensional perception of sound. It is defined as "[…] that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar" [ANSI 1973]. *Timbre* is what enables us to distinguish the sound of A4 in mezzoforte on a violin from that on a piano. Ever since Helmholtz wrote about timbre perception in his book [Helmholtz 1877], there have been many attempts to define *timbre space* that models the human perception of timbre [Grey 1977] [Grey & Gordon 1978] [Krumhansl 1989] [Iverson & Krumhansl 1993] [Kimphoff et al. 1994] [McAdams et al. 1995] [Lakatos 2000] [Caclin et al. 2005] [McAdams et al. 2006] using Multidimensional Scaling (MDS) [Borg & Groenen 1997]. Each work lists the most important acoustic correlates according to their analysis, some of which are common across most works. However, currently there is no unified model that explains the human timbre perception across *all* instrument families.

Rioux and his colleagues at IRCAM collected timbre-related descriptors that are used in music research [Rioux et al. 2006]. It turned out that there are more than 70 descriptors related to timbre perception, but very little systematic evaluation has been done on those descriptors [Giordano & McAdams 2009]. This paper aims at a performance evaluation of those descriptors in two applications of musical instrument classification. One is the instrument family classification, which is to assign instrument sounds according to the instrument family (e.g., string, wind or percussion). The other is the excitation classification problem in the wind instrument family. Same statistical techniques – K-means clustering and correlation analysis – were used for those tasks.

2. ANALYSIS

2. 1. DATA

Six well-known sets of stimuli were used for this project [Grey 1977] [Grey & Gordon 1978] [Krumhansl 1989] [Iverson & Krumhansl 1993] [McAdams et al. 1995] [Lakatos 2000]. Some of these sounds were recordings of real instrument sounds and some others were generated from synthesis. In total, there were 135 instrument sounds, from which 114 stimuli were selected after selecting only the non-modified and non-hybrid stimuli that were studied before. The list of 135 stimuli as well as the 114 chosen stimuli can be seen on Table A1 in Appendix I.

2. 2. DESCRIPTORS & FEATURE VALUES

Rioux et al. lists more than 70 descriptors that are related to timbre perception [Rioux et al. 2006]. A few of them are very similar to others, so the list came down to 70 descriptors after removing those rather redundant ones. Detailed information on each of the descriptors is listed on Table A2 in Appendix I.

The feature values for statistical analyses were obtained by computing each timbre descriptor value on each of the 114 sound stimuli. Therefore the input data for analysis was in the form of a matrix, with 114 rows and 70 columns. The feature values were not normalized.

2. 3. ANALYSIS

There are four research questions that this paper addresses. First, what are the relationships among the descriptors? Second, are there better or worse descriptors in the classification of musical instrument family? Third, is any stimulus consistently harder to correctly classify than others? And fourth, are there better or worse descriptors in the classification of excitation method in the wind instrument family? The following subsections describe the analyses and their interpretations for these questions. All analyses were done in SPSS and MATLAB.

2. 3. 1. Relationships among descriptors

Table 1. The list of 70 descriptors in 9 clusters and the representative(s) of each cluster

CLUSTER	DESCRIPTORS
1. Spectral Slope	DECI, DECIDB, **DECS**
2. Spectral Centroid	CGSILO, CGSMAX, **CGSRMS**, CGSMOY, CGSH, CGSI, CGSB, CGSA, **CGS**, CGSB2, CGSC
3. Spectral Flux	FLMAX, **FLRMS**, FLMOY, FLI
4. Spectral Spread (STD)	STDILO, STDI, STDH, **STDMOY**, STDRMS, STDMAX, STDB, STD, **STDB2**, STDC, STDA
5. Spectral Deviation	DEVMAX, DEVRMS, **DEVMOY**, DEVIDB, DEVS
6. Spectral Shape	SKEW, **KURT**, IPH, **VSRATE**, SLOPE, STDIDB, CGSIDB
7. Fluctuation & Roughness	MAGCO, ROUGH, **FLUC**, **MAXIMUM**
8. RMS Power & Energy	LDBA, **LDBB**, LDBC, LDB, NRGB, NRGI, NRGH, CGT, STDT, **ED**, ITMPN1, MIX, ACUM
9. Attack Time	DEVI, HAC, ITMPN3, **ITMPN2**, ITMPN4, VSPH, VSPC, LTMM, LTMLM, LTMR, LTMLR, LAT

The first question this paper addresses is the relationships among timbre descriptors. For that purpose, Correlation Analysis was used on the descriptors that belong to each of the nine clusters specified by Rioux et al. [Rioux et al. 2006] For each cluster, correlation coefficients were calculated on every pair of feature values belonging to that cluster and either one or two descriptors are chosen that have the highest correlation with other descriptors, hence best-representing their cluster. A few clusters seemed to be further divided into two subclusters, where there are very high intra-correlations among the descriptors within each subcluster, but little inter-subcluster correlations. Two descriptors were chosen in this case to represent each of the subclusters. Table 1 shows the representatives of the nine clusters, which are in bold fonts.

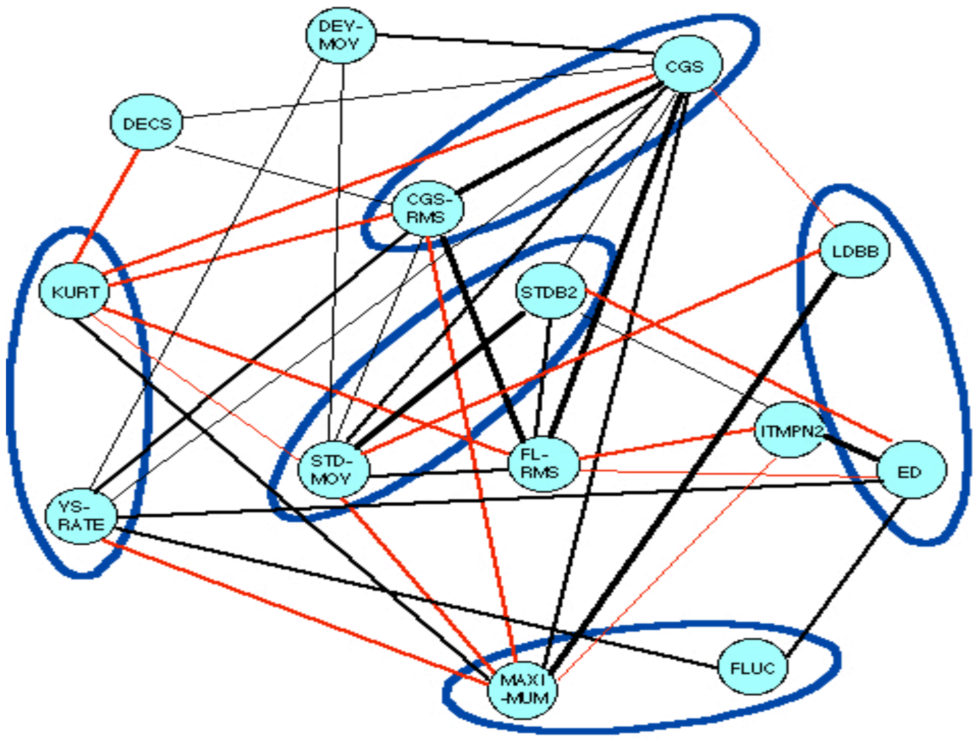

Figure 1. Correlation diagram of the 14 representative descriptors

After narrowing down to 14 representatives (from 70 descriptors), another set of correlation analysis was carried out to figure out the relationships between every possible pair of the 14 descriptors. This gives us a rough idea of the relationships among the descriptors. Figure 1 illustrates the analysis result. The black lines represent positive correlations and the red ones negative correlations. The length of a line does not have any meaning. What is important is the thickness of the lines – a thicker line represents a bigger degree (either positive or negative) of correlation between two descriptors. Some descriptors are grouped into a same group by a thick dark blue circle, which means that they belong to the same cluster. With the cases of CGS – CGSRMS and STDMOY – STDB2, we can see that there are rather thick lines connecting the pairs of descriptors, representing that these values are highly correlated. With the other three cases, we can see that the descriptors within each of the circles do not have any line connecting them directly. This is because the descriptors are not related measures of each other even though they belong to the same cluster from hierarchical cluster analysis. All the other single nodes represent four clusters that are represented by each of them. We can imagine a cloud of unspecified descriptors around each representative one on figure 1, which will give us an insight on how they are inter-related.

2. 3. 2. Classification of musical instrument family

K-means clustering was applied on every column of the input matrix (corresponding to each of the descriptors) for musical instrument family classification. Each stimulus belonged to one of the three instrument groups (wind, string and percussion) and the clustering output from SPSS was recorded for performance evaluation. The performance was then evaluated by counting the number of correctly classified stimuli, which is depicted in figure 2. The best performance was 61% with each of MAGCO (from the fluctuation & roughness cluster) and ITMPN2 (from the attack time cluster) and the worst was 34 % with each of ACUM and ROUGH (both from the fluctuation & roughness cluster).

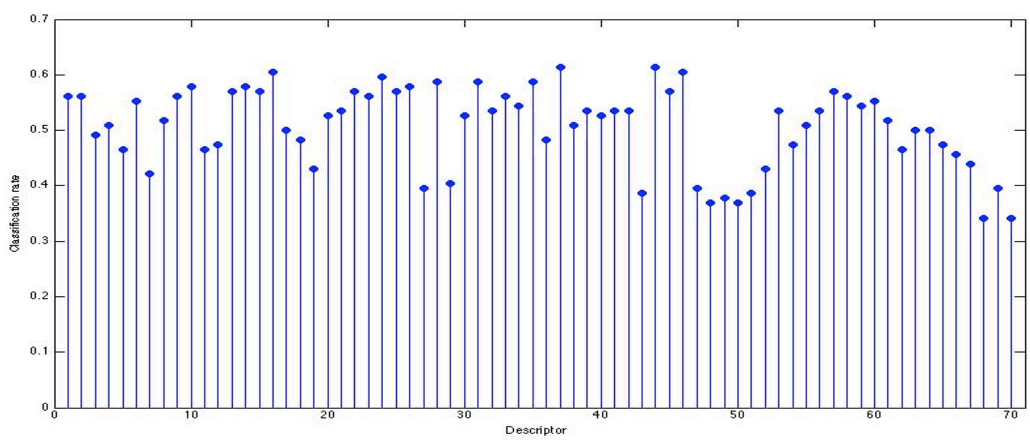

Figure 2. The classification rate per descriptor from K-means clustering

Figure 3 shows the classification rate per descriptor for each of the nine groups. From these graphs, we can see that spectral centroid, spectral flux and spectral deviation are more stable in performance, while roughness or attack time is not. Note that this figure does not say that the descriptors in roughness or attack time clusters are unstable in performance in general. They could still be consistently better than others in other types of classifications, even though they were not as consistent in the task of musical instrument family classification.

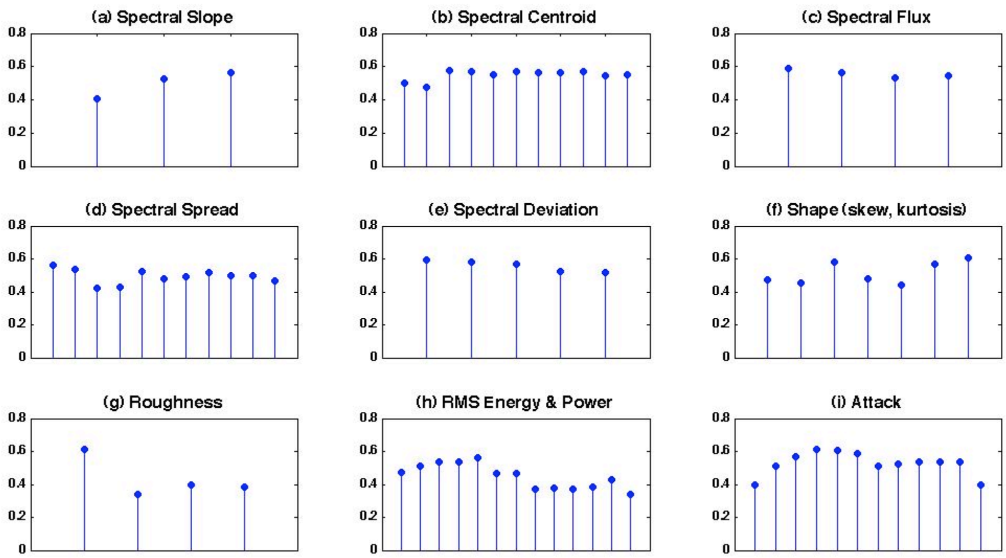

Figure 3. The classification rate per descriptor in nine clusters

2. 3. 3. Performance Evaluation of Stimuli

The question of "are there any stimuli that are consistently better or worse than others in the instrument family classification?" came about during the interpretations of the K-means analysis result in figure 2. Further, I noticed that the three sets of stimuli from Iverson & Krumhansl [Iverson & Krumhansl 1993] seem to follow a very similar pattern and this was worth a further investigation. Figure 4 shows the three sets of stimuli in question. The X-axis lists the instruments corresponding to each stimulus and the Y-axis shows the best classification rate per stimulus across all 70 descriptors. In this figure, we can see that the Remainder curve (red) closely follows the Whole note curve (black) while the Onset only curve seems to show a similar pattern with an offset in the means. There are only a couple of places where all three curves coincide – MTP, Pf and TB. This seems to make sense considering that Pf and TB have relatively strong onsets in comparison with other stimuli, therefore the onset portion of the instrument sounds carries as much information as the remainder (or the whole tone). In the case of MTP, the *mute* applied to the trumpet changes the timbre, which seems to affect the onset portion of the sound so that all three cases (onset only, remainder or whole note) show the same performance. The exact nature of the change of timbre cause by the mute is beyond the scope of this study.

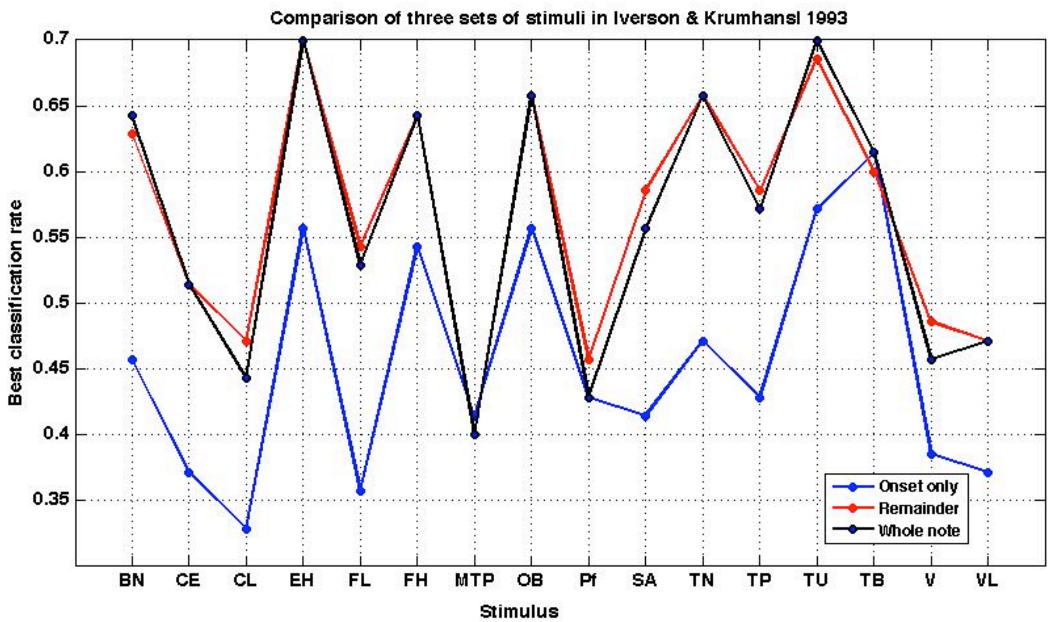

Figure 4. Comparison of three sets of stimuli from [Iverson & Krumhansl 1993] (BN = Bassoon, CE = Cello, CL = Clarinet, EH = English Horn, FL = Flute, FH = French Horn, MTP = Muted Trumpet, OB = Oboe, Pf = Piano, SA = Saxophone, TN = Trombone, TP = Trumpet, TU = Tuba, TB = Tubular Bells, V = Vibraphone, VL = Violin)

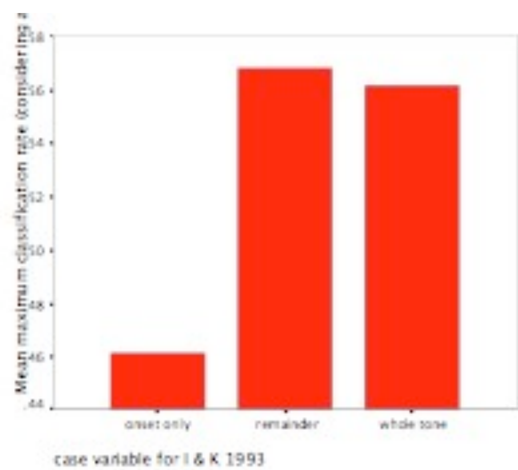

Figure 5. F-test result on the three sets of stimuli from [Iverson & Krumhansl 1993]

With all the other stimuli (than MTP, Pf and TB), there seems to be a consistent offset between the onset-only case and the other two cases. An F-test on SPSS reveals that there is indeed a significant difference among the three means and a further post hoc test (Tukey's) verifies that the onset only condition is significantly different from the other two and those two (remainder and whole tone) are not significantly different. This is illustrated in figure 5.

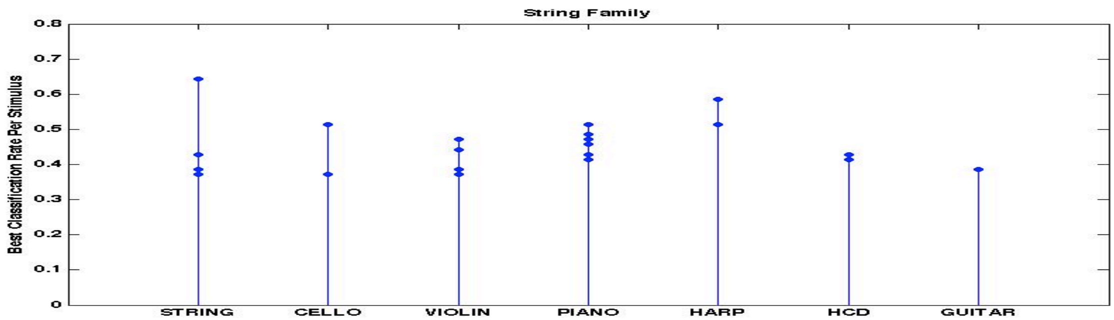

Figure 6. The classification of each stimulus in the string family (HCD = Harpsichord)

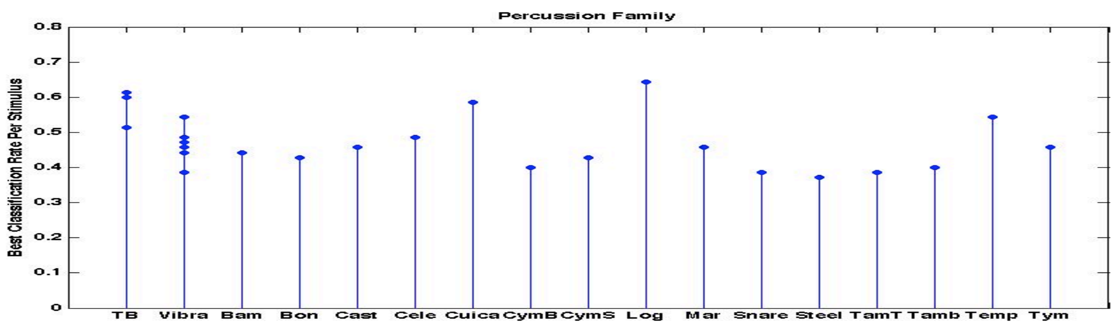

Figure 7. The classification of each stimulus in the percussion family (TB = Tubular bells, Bam = Bamboo chimes, Bon = Bongo drum, Cast = Castanets, Cele = Celesta, CymB = Bowed Symbal, CymS = Struck Symbal, Log = Log drum, Mar = Marimba, Snare = Snare drum, TamT = TamTam, Tamb = Tambourine, Temp = Temple block, Tym = Tympani)

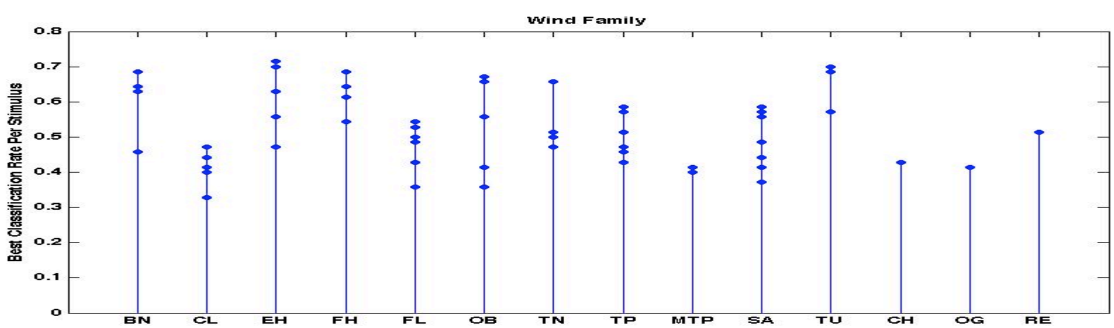

Figure 8. The classification of each stimulus in the wind family (BN = Bassoon, CL = Clarinet, EH = English Horn, FH = French Horn, OB = Oboe, TN = Trombone, TP = Trumpet, MTP = Muted Trumpet, SA = Saxophone, TU = Tuba, CH = Crumhorn Tenor, OG = Baroque Organ, RE = Baroque Recorder)

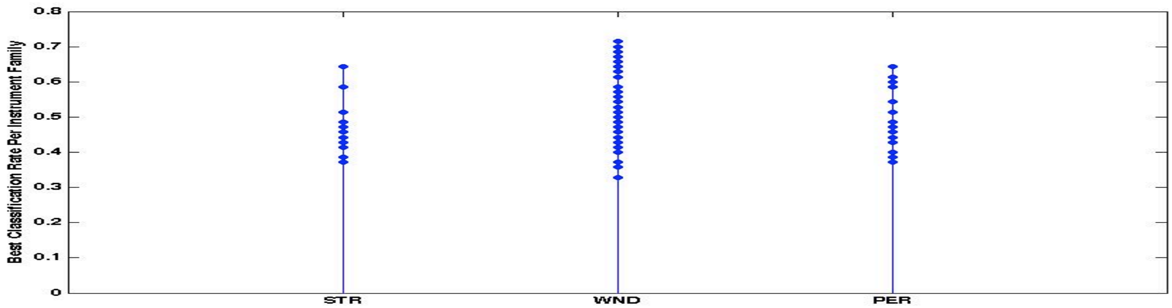

Figure 9. The classification of all stimuli in all three instrument families (STR = string, WND = wind, PER = percussion)

Note that in figure 4, CL is consistently worse than the BN. This motivated a close look at the classification performance in each instrument family, which are shown in figures 6 through 8. The x-axis is various stimuli and the y-axis is the best classification rate of one stimulus across all 70 descriptors. In figure 6, there are a few 'string', cello and violin points that seem to show lower than the average classification rate of the string family. In figure 7 with the percussion family, it is hard to say whether there are certain instruments with consistently lower classification rate since most of the stimuli were just one sample per instrument. In the case of the wind family shown in figure 8, CL, FL and MTP seem to show a consistently lower classification rate than the BN, TU or FH. One possible explanation may be that some instruments (such as BN, TU or FH) have a better classification rate because they are closer to the center of the 'wind' cluster from the K-means clustering while some other instruments (such as CL, FL or MTP) are farther away, thus with a bigger chance of misclassification. More data are required for a verification of this conjecture.

Figure 9 shows the classification performance of all instruments in three families in one graph. It seems that each family (e.g., wind) follows a normal distribution, which is as expected from the Central Limit Theorem. However, there are a couple points in a small group with the classification rate much smaller than the mean in each instrument family, which may support the conjecture that there may be some instruments that are consistently harder to classify correctly than others. Unfortunately we do not have enough number of data points to either prove or disprove this conjecture. This will have to be verified in the future.

2. 3. 4. Classification of excitation type in the wind instrument family

The last research question this paper addresses is the classification of excitation type within the wind instrument family. There are four excitation types, which are listed on Table 2.

Table 2. Four Excitation Types in Wind Instrument Family and the Instruments

Excitation	Instruments
Airjet	Flute, Baroque Recorder
Single Reed	Clarinet, Saxophone
Double Reed	Oboe, Bassoon, Crumhorn, English Horn
Lip Reed	(Muted) Trumpet, Trombone, French Horn, Tuba

As in the instrument family classification, K-means (with K=4 this time) clustering was used on the 14 selected descriptors as determined from the Correlation Analysis. Among the 65 wind stimuli considered for the instrument family classification task, the baroque organ stimulus was excluded since the type of the organ pipe was not specified (which determines the exact excitation type of the stimulus). The clustering result was compared with the true classification and the correct numbers of classifications were counted for each descriptor. Figure 10 shows the performance of each descriptor.

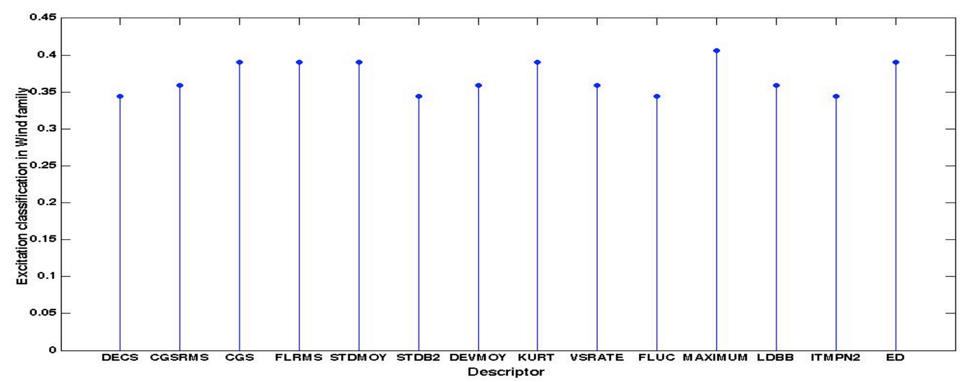

Figure 10. The excitation classification performance of 14 representative descriptors

Even though the most representative descriptors were used for the classification task, the performance was not very good – maximum of 41 % correct. Unlike in the instrument family classification case, the MAXIMUM descriptor (in the fluctuations & roughness cluster) showed the best performance though all of them were in a similar range of performance (34 – 41 % correct). This is certainly not as good as what was hoped for, but it is still not too bad at all considering the poor performance in human classification of instrument's excitation.

3. CONCLUSION & FUTURE WORKS

This paper considered 70 timbre-related descriptors for the musical instrument classification applications. Correlation analysis was applied first to figure out the relationships among the clusters of descriptors. 14 descriptors were chosen to be representatives of the nine clusters and the relationships among them were studied, which will be useful in various timbre-related tasks. K-means clustering analysis was used for the actual classifications of musical instrument family as well as the excitation types within the wind instrument family.

The excitation type classification was not very successful, although it is by nature a harder problem than the musical instrument family classification. The musical instrument family classification showed a better performance in general, achieving over 60 % of correct classification with one descriptor. The next step will be finding an efficient way to combine two or three descriptors for a higher classification rate.

Even though some clusters showed a better and more reliable performance in musical instrument family classification, that was not necessarily the case with the excitation type classification. This seems to mean that the best or the worst descriptor for one task will not necessarily show the same performance in another task. It may be worth an investigation to narrow down the list of "more efficient" descriptors for a few popular tasks.

One thing I had considered but did not do for this project was normalizing the feature values. This may provide a better way of combining descriptors with desired weights. Currently, some values are very small (e.g., STDB) while some others are very big (e.g., MAXIMUM). If STDB is combined with MAXIMUM without any normalization, the performance will be dominated by the performance of MAXIMUM, which will not be desirable. After an efficient combination of descriptors is found, the next step will be an investigation of efficient algorithms, possibly based on machine learning concepts.

BIBLIOGRAPHY

[ANSI 1973] American National Standard Institute. 1973. American national psychoacoustical terminology. S3.20. New York: American Standards Association.

[Borg & Groenen 1997] Borg, I., & Groenen, P. 1997. *Modern multidimensional scaling.* New York, NY: Springer-Verlag.

[Caclin et al. 2005] Caclin, A., McAdams, S., Smith, B.K., & Winsberg, S. 2005. Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *Journal of Acoustical society of America*, *118(1),* 471—82.

[Clark et al. 1964] Clark, M., Robertson, P., & Luce, D.A. 1964 A preliminary experiment on the perceptual basis for musical instrument families. *Journal of the Audio Engineering Society, 12*, 199—203.

[Grey 1977] Grey, J.M. 1977. Multidimensional perceptual scaling of musical timbres. *Journal of Acoustical Society of America, 61*, 1270—1277.

[Grey & Gordon 1978] Grey, J. M., & Gordon, J. W. 1978. Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America, 63*, 1493—1500.

[Giordano & McAdams 2009] Giordano, B., & McAdams, S. 2009. Mechanics of the sound source and the perception of musical timbre: Evidence from previous studies. *Submitted.*

[Handel & Erickson 2004] Handel, S., & Erickson, M.L. 2004. Sound source identification: the possible role of timbre transformations. *Music Perception, 21*(4), 587—610.

[Helmholtz 1877] Helmholtz, H.L.F. 1877. *On the Sensations of Tone as a Psychological Basis for the Theory of Music*. New York, NY: Dover (English translation in 1954).

[Herrera-Boyer et al. 2003] Herrera-Boyer, P., Peeters, G., Dubnov, S. 2003. Automatic Classification of Musical Instrument Sounds. *Journal of New Music Research*, Vol. 32, No. 1, 3—21.

[Iverson & Krumhansl 1993] Iverson, P., & Krumhansl, C.L. 1993. Isolating the dynamic attributes of musical timbre. *Journal of Acoustical Society of America, 94,* 2593—2603.

[Krumhansl 1989] Krumhansl, C.L. Why is musical timbre so hard to understand? In *Structure and Perception of Electroacoustics Sound and Music*, edited by S. Nielzen and O. Olsson (Excerpta Medica, Amsterdam), 44—53.

[Krimphoff et al. 1994] Krimphoff, J., McAdams, S., & Windberg, S. 1994. Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique. *Journal de Physique, 4*, 625—628.

[Krumhansl 1989] Krumhansl, C.L. 1989. Why is musical timbre so hard to understand? In: S. Nielzenand & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43—53). Amsterdam: Elsevier.

[Lakatos 2000] Lakatos, S. 2000. A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*. 62 (7), 1426—1439, 2000.

[Livshin & Rodet 2004] Livshin, A., & Rodet, X. 2004. Musical Instrument Identification in Continuous Recordings. In *Proceedings of the 7th Conference on Digital Audio Effects (DAFx'04), Naples, Italy, October 5 - 8, 2004*, 222—227.

[Martin & Kim 1998] Martin, K.D., & Kim, Y.E., 1998. Musical instrument identification: A pattern-recognition approach. In *Proceedings of the 136th meeting of the Acoustical Society of America.*

[McAdams et al. 1995] McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. 1995. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychol Res, 58*, 177—192.

[McAdams et al. 2006] McAdams, S., Giordano, B., Susini, P., Peeters, G. 2006. A meta-analysis of acoustic correlates of timbre dimensions. *The 9th DAFx* keynote slides (http://www.dafx.ca/slides/keynote3.pdf).

[Rioux et al. 2006] Rioux, V., McAdams, S., Susini, P., Peeters, G. 2006. Psycho-acoustic Timbre Descriptors, IRCAM WP2.1.5. (http://recherche.ircam.fr/projects/cuidado/wg/management/deliverables/IRCAM_WP215_M2.pdf)

[Saldanha & Corso 1964] Saldanha, E.L., & Corso, J.F. 1964. Timbre cues and the identification of musical instruments. *The Journal of the Acoustical Society of America, 36*(10), 2021—2026.

[Wieczorkowska & Kolczynska 2008] Wieczorkowska, A., & Kolczynska, E. 2008. Quality of Musical Instrument Sound Identification for Various Levels of Accompanying Sounds. In *Lecture Notes in Computer Science*. Springer-Verlag Berlin Heidelberg, 2008, 93—103.

APPENDIX I

TABLE A1. The list of 135 stimuli and their references

Reference	Number	Abbrev.	Description	Family	Note
[Grey 1977] [Grey & Gordon 1978]	1	BN	Bassoon	Wind	
	2	BNt	Modified bassoon		Modified
	3	C1	Clarinet 1	Wind	
	4	C2	Clarinet 2	Wind	
	5	C2t	Modified Clarinet 2		Modified
	6	EH	English Horn	Wind	
	7	FH	French Horn	Wind	
	8	FHt	Modified French Horn	Wind	Modified
	9	FL	Flute	Wind	
	10	O1	Oboe1	Wind	
	11	O1t	Modified Oboe 1		Modified
	12	O2	Oboe 2	Wind	
	13	S1	String 1	String	
	14	S1t	Modified String 1		Modified
	15	S2	String 2	String	
	16	S2t	Modified String 2		Modified
	17	S3	String 3	String	
	18	TM	Trombone	Wind	
	19	TMt	Modified Trombone		Modified
	20	TP	Trumpet	Wind	
	21	TPt	Modified Trumpet		Modified
	22	X1	Saxophone 1	Wind	
	23	X2	Saxophone 2	Wind	
	24	X3	Saxophone 3	Wind	
[Iverson & Krumhansl 1993]	25	OBassoon	Bassoon (onset only)	Wind	
	26	OCello	Cello (onset only)	String	
	27	OClarinet	Clarinet (onset only)	Wind	
	28	OEnglishHorn	EnglishHorn (onset only)	Wind	
	29	OFlute	Flute (onset only)	Wind	
	30	OFrenchHorn	FrenchHorn (onset only)	Wind	
	31	OMutedTrumpet	MutedTrumpet (onset only)	Wind	
	32	OOboe	Oboe (onset only)	Wind	
	33	OPiano	Piano (onset only)	String	
	34	OSaxophone	Saxophone (onset only)	Wind	

	35	OTrombone	Trombone (onset only)	Wind	
	36	OTrumpet	Trumpet (onset only)	Wind	
	37	OTuba	Tuba (onset only)	Wind	
	38	OTubularBellsOR	TubularBells (onset only)	Percussion	
	39	OVibraphone	Vibraphone (onset only)	Percussion	
	40	OViolin	Violin (onset only)	String	
	41	RBassoon	Bassoon (remainder)	Wind	
	42	RCello	Cello (remainder)	String	
	43	RClarinet	Clarinet (remainder)	Wind	
	44	REnglishHorn	EnglishHorn (remainder)	Wind	
	45	RFlute	Flute (remainder)	Wind	
	46	RFrenchHorn	FrenchHorn (remainder)	Wind	
	47	RMutedTrumpet	MutedTrumpet (remainder)	Wind	
	48	ROboe	Oboe (remainder)	Wind	
	49	RPiano	Piano (remainder)	String	
	50	RSaxophone	Saxophone (remainder)	Wind	
	51	RTrombone	Trombone (remainder)	Wind	
	52	RTrumpet	Trumpet (remainder)	Wind	
	53	RTuba	Tuba (remainder)	Wind	
	54	RTubularBellsOR	TubularBells (remainder)	Percussion	
	55	RVibraphone	Vibraphone (remainder)	Percussion	
	56	RViolin	Violin (remainder)	String	
	57	WBassoon	Bassoon (whole tone)	Wind	
	58	WCello	Cello (whole tone)	String	
	59	WClarinet	Clarinet (whole tone)	Wind	
	60	WEnglishHorn	EnglishHorn (whole tone)	Wind	
	61	WFlute	Flute (whole tone)	Wind	
	62	WFrenchHorn	FrenchHorn (whole tone)	Wind	
	63	WMutedTrumpet	MutedTrumpet (whole tone)	Wind	
	64	WOboe	Oboe (whole tone)	Wind	
	65	WPiano	Piano (whole tone)	String	
	66	WSaxophone	Saxophone (whole tone)	Wind	
	67	WTrombone	Trombone (whole tone)	Wind	
	68	WTrumpet	Trumpet (whole tone)	Wind	
	69	WTuba	Tuba (whole tone)	Wind	
	70	WTubularBellsOR	TubularBells (whole tone)	Percussion	
	71	WVibraphone	Vibraphone (whole tone)	Percussion	
	72	WViolin	Violin (whole tone)	String	

[Lakatos 2000]	73	BambooChimesHit	BambooChimesHit	Percussion	
	74	BongoSmall	BongoSmall	Percussion	
	75	Castanets	Castanets	Percussion	
	76	Celesta	Celesta	Percussion	
	77	ClarinetBb	ClarinetBb	Wind	
	78	CrumhornTenor	CrumhornTenor	Wind	
	79	Cuica	Cuica	Percussion	
	80	CymbalBowed	CymbalBowed	Percussion	
	81	CymbalStruck	CymbalStruck	Percussion	
	82	DoubleBassPizz	DoubleBassPizz	String	Not tested
	83	EnglishHorn	EnglishHorn	Wind	
	84	FluteFlutterTongued	FluteFlutterTongued	Wind	
	85	FluteNoVibe	FluteNoVibe	Wind	
	86	FrenchHorn	FrenchHorn	Wind	
	87	Harp	Harp	String	
	88	Harpsichord	Harpsichord	String	
	89	LogDrum	LogDrum	Percussion	
	90	Marimba	Marimba	Percussion	
	91	OrganBaroquePlenum	OrganBaroquePlenum	Wind	
	92	Piano	Piano	String	
	93	RecorderBaroqueTenor	RecorderBaroqueTenor	Wind	
	94	SaxAlto	SaxAlto	Wind	
	95	SaxTenorGrowls	SaxTenorGrowls	Wind	
	96	Snare	Snare	Percussion	
	97	SteelDrum	SteelDrum	Percussion	
	98	TamTam	TamTam	Percussion	
	99	TambourinePop	TambourinePop	Percussion	
	100	TempleBlock	TempleBlock	Percussion	
	101	TrumpetBbHard	TrumpetBbHard	Wind	
	102	TrumpetMuted	TrumpetMuted	Wind	
	103	TubularBells	TubularBells	Percussion	
	104	Tympani	Tympani	Percussion	
	105	VibraphoneBowed	VibraphoneBowed	Percussion	
	106	VibraphoneHardMallet	VibraphoneHardMallet	Percussion	
	107	ViolinMartele	ViolinMartele	String	
	108	ViolinNoVibe	ViolinNoVibe	String	
[Krumhansl 1989]	109	bsn	bassoon	Wind	
	110	can	cor anglais (tenor oboe)	Wind	

[McAdams et al. 1995]	111	clarinette	clarinette	Wind	Not tested
	112	clavecin	clavecin	String	Not tested
	113	cnt	clarinet	Wind	
	114	gtn	guitarnet (guitar / clarinet)	N/A	Hybrid
	115	gtr	guitar	String	
	116	hcd	harpsichord	String	
	117	hrn	french horn	Wind	
	118	hrp	harp	String	
	119	obc	obochord (oboe / harpsichord)	N/A	Hybrid
	120	obo	Oboe	Wind	
	121	ols	oboleste (oboe / celesta)	N/A	Hybrid
	122	piano	piano	String	Not teseted
	123	pianofrot	pianofrot	?	Not tested
	124	pno	piano	String	
	125	pob	pianobow (bowed piano	String	
	126	sno	striano (bowed string / piano)	N/A	Hybrid
	127	spo	sampled piano	String	
	128	stg	bowed string	String	
	129	tbn	trombone	Wind	
	130	tpr	trumpar (trumpet / guitar)	N/A	Hybrid
	131	tpt	trumpet	Wind	
	132	trompette	trompette	Wind	Not tested
	133	vbn	vibrone (vibraphone / trombone)	N/A	Hybrid
	134	vbs	vibraphone	Percussion	
	135	pbo	?	?	Not tested

TABLE A2. The list of 70 timbre descriptors and their references

Group	Subgroup	Name	Description
Harmonic [Peeters 2000]	Spectrum	NRGB	energy
		CGSB	spectral centroid (global mean spec) [cgsb]
		STDB	spectral variation
		VSPC	spectral centroid (global mean spec) [vspc]
	Harmonic	NRGH	spectral energy
		CGSH	spectral centroid
		STDH	spectral std
		DEVS	spec deviation (of the harmonic computed from the global mean spectrum)
		DECS	spec slope
		IPH	something related to a group of spectral shape descriptors (*)
		NRGI	mean of the instantaneous energy
		CGSMAX	spectral centroid computed on the vector composed of the maximum amplitude [lin] of each harmonic over time
		CGSMOY	spec centroid computed on the vector composed of the mean amplitude [lin] of each harmonic over time
		CGSRMS	spec centroid computed on the vector composed of the rms amplitude [lin] of each harmonic over time
		CGSI	mean of the instantaneous spec centroid [amp lin, freq lin]
		CGSIDB	mean of the instantaneous spec centroid [amp dB, freq lin]
		CGSILO	mean of the instantaneous spec centroid [amp lin, freq log]
		STDMAX	spectral std computed on the vector composed of the maximum amplitude [lin] of each harmonic over time
		STDMOY	spectral std computed on the vector composed of the mean amplitude [lin] of each harmonic over time
		STDRMS	spectral std computed on the vector composed of the rms amplitude [lin] of each harmonic over time
		STDI	mean of the instantaneous spec std [amp lin, freq lin]
		STDIDB	mean of the instantaneous spec std [amp dB, freq lin]
		STDILO	mean of the instantaneous spec std [amp lin, freq log]
Harmonic2 [Peeters 2000]	Harmonic	DEVMAX	spectral std computed on the vector composed of the maximum of amplitude [dB] of each harmonic over time
		DEVMOY	spectral std computed on the vector composed of the mean of

			amplitude [dB] of each harmonic over time
		DEVRMS	harmonic - spectral std computed on the vector composed of the rms of amplitude [dB] of each harmonic over time
		DEVI	mean of the instantaneous spec deviation [amp lin]
		DEVIDB	mean of the instantaneous spec deviation [amp dB]
		DECI	mean of the instantaneous spec slope [amp lin]
		DECIDB	mean of the instantaneous spec slope [amp dB]
		FLMAX	spec flux using instantaneous centroid and cgsmax
		FLMOY	spec flux using instantaneous centroid and cgsmoy
		FLRMS	spec flux using instantaneous centroid and cgsrms
		FLI	spec flux using instantaneous centroid and cgsi
		VSPH	harmonic spectral deviation
		VSRATE	speed of variation of the spectrum
		MAGCO	sum of the variations of the instantaneous harmonic from global mean harmonics
		HAC	harmonic attack coherence
	Envelope	LTMR	log-attack time from [rms]
		LTMM	log-attack time from [max]
		LTMLR	log-attack time from [smoothed rms]
		LTMLM	log-attack time from [smoothed max]
		ITMPN1	effective duration
		ITMPN2	effective duration [norm by file length]
		ITMPN3	effective duration [norm by file length and f0]
		ITMPN4	effective duration [norm by file length and T]
Percussive [Lakatos 2000]		LAT	log-attack time
		CGT	temporal centroid
		STDT	temporal std
		ED	effective duration
		MAXIMUM	maximum value
		MIX	ed*cgt
		LDB	rms value of the power spectrum
		LDBA	rms value of the power spectrum [amp weighting dbA]
		LDBB	rms value of the power spectrum [amp weighting dbB]
		LDBC	rms value of the power spectrum [amp weighting dbC]
		CGS	spec centroid of the power spec

	CGSA	spec centroid of the power spec [amp weighting dbA]
	CGSB2	spec centroid of the power spec [amp weighting dbB]
	CGSC	spec centroid of the power spec [amp weighting dbC]
	STD	spec std of the power spec
	STDA	spec std of the power spec [amp weighting dbA]
	STDB2	spec std of the power spec [amp weighting dbB]
	STDC	spec std of the power spec [amp weighting dbC]
	SKEW	skewness of the power spec
	KURT	kurtosis of the power spec
	SLOPE	slope of the power spec
Other	ACUM	sharpness [Bismarck 74]
	FLUC	fluctuation strength (in vacil) [Susini 00]
	ROUGH	roughness (in aspers) [Aures 85]