CATEGORY-LEVEL IDENTIFICATION OF NON-REGISTERED MUSICAL INSTRUMENT SOUNDS

Tetsuro Kitahara,† *Masataka Goto,*‡ and *Hiroshi G. Okuno*†

†Dept. of Intelligence Science and Technology
Graduate School of Informatics, Kyoto University
Sakyo-ku, Kyoto 606-8501, Japan

‡National Institute of Advanced
Industrial Science and Technology
Tsukuba, Ibaraki 305-8568, Japan

kitahara@kuis.kyoto-u.ac.jp   m.goto@aist.go.jp   okuno@i.kyoto-u.ac.jp

ABSTRACT

This paper describes a method that identifies sounds of *non-registered musical instruments* (*i.e.*, musical instruments that are not contained in the training data) at a category level. Although the problem of how to deal with non-registered musical instruments is essential in musical instrument identification, it has not been dealt with in previous studies. Our method solves this problem by distinguishing between registered and non-registered instruments and identifying the category name of the non-registered instruments. When a given sound is registered, its instrument name, *e.g.* violin, is identified. Even if it is not registered, its category name, *e.g.* strings, can be identified. The important issue in achieving such identification is to adopt a musical instrument hierarchy reflecting the acoustical similarity. We present a method for acquiring such a hierarchy from a musical instrument sound database. Experimental results show that around 77% of non-registered instrument sounds, on average, were correctly identified at the category level.

1. INTRODUCTION

The increasing amount of musical audio signals on the Internet and in a personal storage requires efficient and universal description of them. MPEG-7 [1], which is a new ISO standard, provides a solution for this music description. The names of musical instruments have an important role as music descriptors because musical pieces are sometimes characterized by what instruments are used. In fact, we use music genres having instrument names, such as "piano sonata" and "string quartet." In addition, when a user wants to search musical pieces of piano solos or string quartets, a retrieval system can use descriptors of musical instrument names. Therefore, musical instrument identification, which aims to obtain the names of the instruments used in musical pieces, has been studied in recent years [2]–[7].

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Grant-in-Aid for Scientific Research (A), No.15200015, the Sound Technology Promotion Foundation, and Informatics Research Center for Development of Knowledge Society Infrastructure (COE program of MEXT, Japan).

We focus on a new problem of identifying *non-registered musical instruments*, that is, identifying musical instruments that are not contained in the training data. Most studies [2]–[7] have used training data containing a limited number of musical instruments and have assumed that all the instruments used in an input were contained in the training data. Because there are numerous kinds of musical instruments in the world, it is impossible to prepare training data that covers all of them. In addition, the recent development of digital audio technology has made it possible to create novel and infinite kinds of original musical sounds (from sounds similar to natural instruments to sounds of instruments that do not actually exist). It is therefore essential to deal with non-registered musical instruments when identifying musical instrument sounds.

In this paper, to solve this problem, we propose category-level identification of the non-registered musical instruments. For example, a musical instrument sound that is similar to a violin and a viola but not the same (for example, a sound made from the two instruments using a synthesizer) is identified as "strings." When humans listen to this sound for the first time, they would think "I do not know this instrument, but it must be a kind of strings." This study aims to achieve such human-like recognition on a computer.

This paper also discusses a *musical instrument hierarchy* (MIH) for this category-level identification. The most important requirement for the MIH in category-level identification is that it should reflect the similarity of timbres (acoustical features). However, MIHs satisfying this requirement have not been reported in the literature. We present a method for automatic acquisition of the MIH based on the acoustical similarity of musical instruments. This acoustical-similarity-based MIH is called *AcoustMIH*.

The rest of this paper is organized as follows: Section 2 presents a method for acquiring AcoustMIH. Section 3 reports experiments on identifying the non-registered musical instruments. Finally, Section 4 concludes this paper.

2. ACOUST-MIH

AcoustMIH is the MIH based on the acoustical similarity for

Table 1. A conventional hierarchy of musical instruments [8].

Higher level	Middle level	Lower level	Musical instruments*
Strings	—	Struck strings	PF
		Plucked strings	CG, UK, AG
		Bowed strings	VN, VL, VC
Winds	Wood winds	Air reeds	PC, FL, RC
		Single reeds	SS, AS, TS, BS, CL
		Double reeds	OB, FG
	Brasses	(Rip reeds)	TR, TB
Percuss.	(omitted)	(omitted)	(omitted)

*Notation of musical instruments is defined in **Table 3**.

category-level musical instrument identification.

The MIH for category-level identification should reflect the timbre similarity. In other words, two instruments that are close on the hierarchy should have similar timbres. However, most of the commonly used hierarchies do not satisfy this requirement. For example, in the hierarchy showed in **Table 1** [8] which is designed based on sounding mechanisms and playing methods of musical instruments, both pianos (PF) and violins (VN) belong to the same category, *strings*, but their timbres are quite different.

In this paper, we present a method for automatic acquisition of AcoustMIH, which satisfies the above requirement, using a large musical instrument sound database. The rest of this section discusses problems, solutions and results of automatic acquisition of AcoustMIH.

2.1. Problems and Our Solutions

One of the most commonly used methods for acquiring a hierarchy from feature vectors is *hierarchical clustering*. Hierarchical clustering first calculates distances between feature vectors in a feature space and then merges the closest pair of feature vectors (or clusters) into a single cluster recursively until all the feature vectors are merged into a single cluster. This method can be applied to acquiring AcoustMIH, but the following two problems make it difficult to obtain reasonable results:

Problem 1 Clustering results depend on a feature space.

Problem 2 If one sound is used as a representative of each musical instrument, the clustering results also depend on this sound. This is because features of musical instrument sounds depend on various factors including pitch and differences of individuals.

In this paper, to solve **Problem 1**, we use the same feature space for both identification and clustering. Since different musical instrument identification methods would have different feature spaces, MIHs appropriate for the identification methods would also be different. Our approach makes it possible to cope with the dependency on identification methods. To solve **Problem 2**, we perform hierarchical clustering on a multivariate normal distribution (MND) of each instrument, which is obtained from a large musical instrument sound database. By using an MND, instead of a single sound, for each instrument, we can obtain the appropriate representative position of the instrument in the feature space.

Table 2. Overview of 129 features used in [7].

(1)	**Spectral features** (40 features)
	e.g., Spectral centroid, relative power of the fundamental component, relative power in odd and even components
(2)	**Temporal features** (35 features)
	e.g., Gradient of a straight line approximating power envelope, average differential of power envelope during onset
(3)	**Modulation features** (32 features)
	e.g., Amplitude and frequency of AM, FM, modulation of spectral centroid and modulation of MFCC
(4)	**Non-harmonic component features** (22 features)
	e.g., Temporal mean of kurtosis of spectral peaks of each harmonic component (Their values decrease as sounds contain more non-harmonic components.)

2.2. The detail of the method

AcoustMIH is acquired by the following three steps:

1. **Feature Extraction**

Features that are the same as those used for identification are extracted. Since we use a musical instrument identification method presented in our previous paper [7], we extract the same features as those used in the paper [7]. Specifically, 129 features listed in **Table 2** are first extracted, and then the dimensionality of the 129-dimensional feature space is reduced by two successive processing: it is reduced to 79 dimensions by principal component analysis (PCA) with the proportion value of 99%, and then is further reduced by linear discriminant analysis (LDA). The feature space is finally reduced to an 18-dimensional one because we deal with 19 instruments.

2. **Calculation of the Mahalanobis Distances**

Once the distribution of each instrument ω_i in the feature space is approximated by an MND, the mean vector $\boldsymbol{\mu}_i$ and the covariance matrix Σ_i of this distribution are calculated. The Mahalanobis distance $D_{\mathrm{M}}(\omega_i, \omega_j)$ of each instrument pair (ω_i, ω_j) $(\omega_i \neq \omega_j)$ is calculated by the following equation:

$$D_{\mathrm{M}}(\omega_i, \omega_j) = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)' \Sigma_{i,j}^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j),$$

where, $\Sigma_{i,j} = (\Sigma_i + \Sigma_j)/2$, and $'$ represents the transposition operator.

3. **Hierarchical Clustering**

Hierarchical clustering is performed using the above Mahalanobis distances. In this paper, we adopted the average-link clustering, which considers the distance between two clusters to be equal to the average distance from any member of one cluster to any member of the other.

Table 3. Contents of the database used in this paper.

Instrument names	Piano (PF), Classical Guitar (CG), Ukulele (UK), Acoustic Guitar (AG), Violin (VN), Viola (VL), Cello (VC), Trumpet (TR), Trombone (TB), Soprano Sax (SS), Alto Sax (AS), Tenor Sax (TS), Baritone Sax (BS), Oboe (OB), Fagotto (FG), Clarinet (CL), Piccolo (PC), Flute (FL), Recorder (RC)
Individuals	3 individuals except TR, OB, FL. TR, OB, FL: 2 individuals.
Intensity	Forte, normal, piano.
Articulation	Normal articulation style only.
Number of tones	PF: 508, CG: 696, UK: 295, AG: 666, VN: 528, VC: 558, TR: 151, TB: 262, SS: 169, AS: 282, TS: 153, BS: 215, OB: 151, FG: 312, CL: 263, PC: 245, FL: 134, RC: 160.

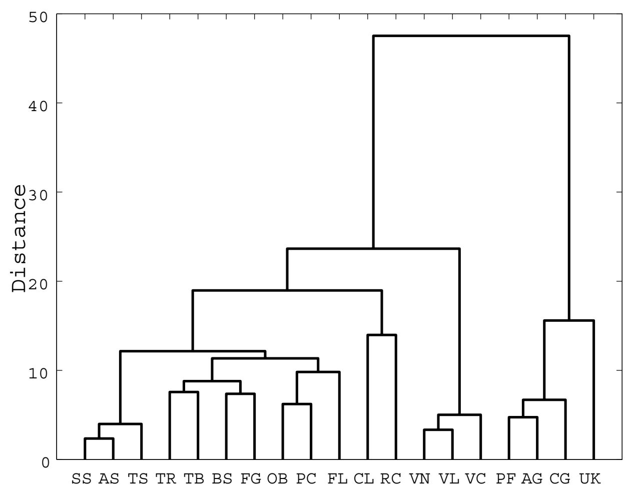

Fig. 1. The musical instrument hierarchy acquired by the proposed method.

2.3. Actual Acquisition of AcoustMIH

We conducted experiments on automatic acquisition of AcoustMIH using a subset of a large musical instrument sound database RWC-MDB-I-2001 [9]. This subset summarized in **Table 3** was selected by the quality of recorded sounds. It consists of 6,247 solo tones of 19 orchestral instruments. All data were sampled at 44.1 kHz with 16 bits.

AcoustMIH, acquired by the proposed method, is shown in **Fig. 1**. We obtained musical instrument categorization by merging musical instruments of which distances in **Fig. 1** are less than a threshold each other into one cluster. Higher, middle and lower levels in **Table 4** show the categorization obtained when the threshold is 30, 20 and 10, respectively.

We conducted preliminary experiments on category-level identification of registered musical instruments. We assigned half the data in **Table 3** to training data and the rest

Table 4. Musical instrument categorization at three different levels obtained by **Fig. 1**.

Higher level	Middle level	Lower level	Musical Instruments
Decayed	—	Ukulele	UK
		Others	PF, CG, AG
Sustained	Strings	—	VN, VL, VC
	Woods	Saxophones	SS, AS, TS
		Clarinet	CL
		Recorder	RC
		Brasses, etc.	TR, TB, BS, FG
		Others	OB, PC, FL

Table 5. Musical instrument sounds used for identification of non-registered musical instruments.

Sound names	Electric Piano (ElecPf), Synth Strings (SynStr), Synth Brass (SynBrs)
Variation	2 variations for each sound name
Velocity	100
Pitch range	C3–C5 (A4=440Hz)

to test data. By using AcoustMIH (**Table 4**, lower level), we attained 90.81% of recognition, while the recognition rate using the conventional one (**Table 1**, lower level) was 88.85%.

3. CATEGORY-LEVEL IDENTIFICATION OF NON-REGISTERED MUSICAL INSTRUMENTS

In this section, we report experiments on category-level identification of non-registered musical instruments using the musical instrument categorization obtained by AcoustMIH. We used the sounds listed in **Table 3** as training data and electric sounds played by a MIDI tone generator (MU2000, Yamaha), listed in **Table 5**, as non-registered musical instrument sounds.

3.1. Category-level Identification of Non-registered Musical Instrument Sounds

The results of identifying the non-registered musical instruments at the category level are shown in **Table 6**. Recognition rates using AcoustMIH (**Table 4**, lower level) were between 75 and 100%. On the other hand, recognition rates by the conventional hierarchy (**Table 1**, lower level) were lower except for synth strings. These results suggest that sounding-mechanism-based categorization is unsuitable for electric sounds, since they do not have sounding mechanisms.

3.2. Determination of Whether Musical Instruments Are Registered or Not

We conducted experiments on determining whether musical instruments are registered or not. This determination is required for flexible musical instrument identification, that is, instrument-name-level identification for registered instruments and category-level identification for non-registered instruments. This is performed by the following steps:

1. Recognize a musical instrument sound at instrument name level;
2. Calculate the Mahalanobis distance from the sound to the distribution of the above result;
3. Judge it to be *registered* if the distance is less than a threshold, or *non-registered* if the distance is not.

To calculate the Mahalanobis distances, a 23-dimensional feature space obtained by PCA was used.

We assigned half the data in **Table 3** to training data, the rest to test data of registered instruments, and all of the data in **Table 5** to test data of non-registered instruments. Experimental results show that the performance of correct determination was 85% when the threshold was 40. The details were omitted because of a lack of space.

3.3. Flexible Musical Instrument Identification

We finally present results of flexible musical instrument identification. The experimental conditions are the same as those in Section 3.2. The experimental results listed in **Table 7** show that our method correctly identified 67% of registered instrument sounds at the instrument name level, 13% of them at the category level, and 77% of non-registered instrument sounds at the category level while distinguishing them from registered sounds. The recognition rate for ElecPf A was poor, because it was not recognized as a non-registered instrument, but as a registered one. Actually, it sounds like a real piano for human listeners.

4. CONCLUSIONS

In this paper, we pointed out a new problem in musical instrument identification, *the problem of non-registered musical instruments*, and proposed category-level identification of the non-registered instruments as a solution. In addition, we acquired a musical instrument hierarchy based on the acoustical similarity for this category-level identification. Experimental results show that non-registered instrument sounds were correctly distinguished from registered one with accuracy of 85% and their category names were identified with accuracy of 77%. Future work will include evaluation on a mixture of sounds and real musical pieces.

Acknowledgments: We thank everyone who has contributed to building and distributing the RWC Music Database (Musical Instrument Sound: RWC-MDB-I-2001) [9]. We also thank Prof. Haruhiro Katayose, Dr. Kunio Kashino and Dr. Kazuhiro Nakadai for their valuable comments.

Table 6. Experimental results of identifying non-registered musical instruments.

	ElecPf		SynStr		SynBrs	
	A	B	A	B	A	B
Conv.	16%	28%	84%	100%	4%	28%
Prop.	100%	92%	84%	100%	76%	100%

Conv.: Using the hierarchy shown in **Table 1**
Prop.: Using the hierarchy shown in **Table 4** (ours)

Table 7. Experimental results on flexible musical instrument identification.

Registered	PF	CG	UK	AG	VN	VL	VC
Correct I	69%	83%	97%	68%	62%	69%	70%
Correct II	17%	12%	0%	14%	14%	11%	10%
Incorrect	14%	5%	3%	17%	24%	20%	20%
	TR	TB	SS	AS	TS	BS	OB
Correct I	64%	63%	47%	40%	30%	49%	48%
Correct II	15%	17%	11%	17%	26%	20%	19%
Incorrect	22%	20%	42%	43%	44%	31%	33%
	FG	CL	PC	FC	RC	Av.	
Correct I	56%	91%	66%	45%	89%	67%	
Correct II	16%	0%	17%	20%	0%	13%	
Incorrect	27%	9%	17%	35%	11%	20%	

Non registered	ElecPf A	ElecPf B	SynStr A	SynStr B	SynBrs A	SynBrs B	Av.
Correct II	44%	76%	88%	100%	60%	96%	77%
Incorrect	56%	24%	12%	0%	40%	4%	33%

Correct I: Correct at instrument name level.
Correct II: Correct at category level while rejecting instrument-name-level results.

5. REFERENCES

[1] http://ipsi.fhg.de/delite/Projects/MPEG7/

[2] K. D. Martin, "Sound-Source Recognition: A Theory and Computational Model," Ph.D. Thesis, MIT, 1999.

[3] A. Eronen and A. Klapuri, "Musical Instrument Recognition using Cepstral Coefficients and Temporal Features," *Proc. ICASSP*, pp.753–756, 2000.

[4] K. Kashino and H. Murase, "A Sound Source Identification System for Ensemble Music based on Template Adaptation and Music Stream Extraction," *Speech Communication*, **27**, pp.337–349, 1999.

[5] I. Fujinaga and K. MacMillan, "Realtime Recognition of Orchestral Instruments," *Proc. ICMC*, pp.141–143, 2000.

[6] A. A. Livshin, G. Peeters and X. Rodet, "Studies and Improvements in Automatic Classification of Musical Sound Samples," *Proc. ICMC*, pp.171–178, 2003.

[7] T. Kitahara, M. Goto and H. G. Okuno, "Musical Instrument Identification based on F0-dependent Multivariate Normal Distribution," *Proc. ICASSP*, **V**, pp.421–424, 2003.

[8] H. F. Olson, "Music, Physics and Engineering", *Dover Publications*, 1966.

[9] M. Goto, H. Hashiguchi, T. Nishimura and R. Oka, "RWC Music Database: Music Genre Database and Musical Instrument Sound Database," *Proc. ISMIR*, pp.229–230, 2003.