Applied Intelligence 23, 267–275, 2005


# Pitch-Dependent Identification of Musical Instrument Sounds*

TETSURO KITAHARA
*Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan*
kitahara@kuis.kyoto-u.ac.jp

MASATAKA GOTO
*"Information and Human Activity", PRESTO, JST/National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan*
m.goto@aist.go.jp

HIROSHI G. OKUNO
*Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan*
okuno@i.kyoto-u.ac.jp

**Abstract.** This paper describes a musical instrument identification method that takes into consideration the *pitch dependency* of timbres of musical instruments. The difficulty in musical instrument identification resides in the pitch dependency of musical instrument sounds, that is, acoustic features of most musical instruments vary according to the pitch (fundamental frequency, F0). To cope with this difficulty, we propose an *F0-dependent multivariate normal distribution*, where each element of the mean vector is represented by a function of F0. Our method first extracts 129 features (e.g., the spectral centroid, the gradient of the straight line approximating the power envelope) from a musical instrument sound and then reduces the dimensionality of the feature space into 18 dimension. In the 18-dimensional feature space, it calculates an *F0-dependent mean function* and an *F0-normalized covariance*, and finally applies the Bayes decision rule. Experimental results of identifying 6,247 solo tones of 19 musical instruments shows that the proposed method improved the recognition rate from 75.73% to 79.73%.

**Keywords:** musical instrument identification, the pitch dependency, fundamental frequency, automatic music transcription, computational auditory scene analysis

## 1. Introduction

Computational auditory scene analysis as well as visual scene analysis is important to augment communication channels between humans and computers and to achieve sophisticated human-computer interaction [8]. However, auditory information has not been used extensively except for speech. In particular, musical instrument identification is an important subtask for the computational auditory scene analysis.

Musical instrument identification is also important from a point of view of application. For example, a user wants to search for certain types of musical pieces, such as piano solos or string quartets, a retrieval system can use the results of musical instrument identification. Furthermore, the results of musical instrument

*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, and Informatics Research Center for Development of Knowledge Society Infrastructure (COE program of MEXT, Japan).

identification will be important cues for reducing ambiguity when automatically transcribing a musical piece played on multiple instruments.

The difficulties in musical instrument identification reside in the fact that most of features of musical sounds depend on some factors including pitch. In particular, timbres of musical instruments are obviously affected by the pitch due to their wide range of pitch.[1] For example, the pitch range of the piano covers over seven octaves.

To attain high performance of musical instrument identification, it is indispensable to cope with this *pitch dependency* of timbre. Most studies on musical instrument identification, however, have not dealt with the pitch dependency [1–5]. Martin used 31 features including spectral and temporal features with hierarchical classification and attained about 70% of identification by the benchmark of 1,023 solo tones played by 14 instruments. He pointed out the importance of the pitch dependency, but left it as future work [5]. Eronen *et al.* used spectral and temporal features as well as cepstral coefficients used by Brown [1] and attained about 80% of identification by the benchmark of 1,498 solo tones played by 30 instruments [2]. They treated the pitch as one element of feature vectors, but did not cope with the pitch dependency. Kashino *et al.* also treated the pitch similarly in their automatic music transcription system [4]. They also coped with the difference of individuals of musical instruments (for example, the difference of sounds of Yamaha's pianos and Boesendolfer's ones) by template adaptation, but did not deal with the pitch dependency [6].

In this paper, to take into consideration the pitch dependency of timbre, we propose a method for modeling the pitch dependency of timbre, *F0-dependent multivariate normal distribution*. Each feature or basic vector of features extracted from musical instrument sounds is represented by the F0-dependent multivariate normal distribution, the mean of which is represented by a function of fundamental frequency (F0). This *F0-dependent mean function* represents the pitch dependency of each feature, while the *F0-normalized covariance* represents the non-pitch dependency. Musical instrument identification is performed both at individual-instrument level and at non-tree category level by a discriminant function based on the Bayes decision rule.

The rest of this paper is organized as follows: Section 2 proposes the F0-dependent multivariate normal distribution, and Section 3 describes a discriminant function based on the Bayes decision rule. Sections 4 and 5 report the experimental results. Finally, Section 6 concludes this paper.

## 2. Musical Instrument Identification using F0-dependent Multivariate Normal Distribution

The key idea of our method is to approximate the pitch dependency of each feature representing timbres of musical instrument sounds as a function of fundamental frequency (F0). An F0-dependent multivariate normal distribution has two parameters: an F0-dependent mean function and an F0-normalized covariance. The former represents the pitch dependency of features and the latter represents the non-pitch dependency. The reason why the mean of a distribution of tone features is approximated as a function of F0 is that tone features at different pitches have different positions (means) of distributions in the feature space. Approximating the mean of the distributions as a function of F0 makes it possible to model how the features vary according to the pitch with a small set of parameters.

### 2.1. *Frequency Analysis*

Given a musical instrument signal, it is first analyzed by the short-time Fourier transform (STFT) with a 4096-point Hanning window for every 10 ms, and spectral peaks are extracted from the power spectrum. Then the harmonic structure and the F0 is obtained from these peaks.

### 2.2. *Feature Extraction*

The following 129 features are extracted:

(1) Spectral Features

- 1 Spectral centroid
- 2 Relative power of the fundamental component
- 3–30 Relative cumulative power of from fundamental to $i$-th components ($i = 2, 3, \ldots, 29$)
- 31 Relative power in odd and even components
- 32–40 The number of the components the duration of which is $p$% longer than the longest duration ($p = 10, 20, \ldots, 90$)

(2) Temporal Features

41 Gradient of the straight line approximating the power envelope

42–58 Average differential of the power envelope during $t$-sec interval from the onset time ($t = 0.15, 0.20, \ldots, 0.95$)

59–75 The ratio of the power at $t$-sec after the onset time

(3) Modulation Features

76 The amplitude of AM

77 The frequency of AM

78 The amplitude of FM

79 The frequency of FM

80 The amplitude of the spectral centroid modulation

81 The frequency of the spectral centroid modulation

82–94 The amplitude of $k$-th MFCC

95–107 The frequency of $k$-th MFCC

(4) Peak Kurtosis Features

108–118 Temporal average of the kurtosis of spectral peaks in the $i$-th component ($i = 1, 2, \ldots, 11$)

119–129 The amplitude of the temporal modulation of the kurtosis of spectral peaks in the $i$-th component ($i = 1, 2, \ldots, 11$)

Whereas the spectral, temporal and modulation features are designed based on previous studies, the peak kurtosis features are original and have not been used in previous studies. The kurtosis of spectral peaks is related to how large non-harmonic components are included. If a sound has small non-harmonic components, spectral peaks will have large kurtosis. If it has large non-harmonic components, they will have small kurtosis (see Fig. 1). These features are therefore used for modeling the degree of incorporation of non-harmonic components, which has not been considered in previous studies.

After the feature extraction, the feature space is standardized and then the dimensionality of it is reduced by two methods: the 129-dimensional feature space is reduced to a 79-dimensional one by principal component analysis (PCA) with the proportion value of 99%, and then is further reduced to a lower-dimensional one by linear discriminant analysis (LDA). In this paper, the feature space is reduced to an 18-dimensional space, since we deal with 19 instruments.

### 2.3. *Parameter Estimation of F0-dependent Multivariate Normal Distribution*

For each instrument $\omega_i$, the parameters of the F0-dependent multivariate normal distribution, the F0-dependent mean function $\boldsymbol{\mu}_i(f)$ and the F0-normalized covariance $\Sigma_i$, are calculated. The F0-dependent mean function is obtained through approximating each element of the feature vectors as a cubic polynomial with the least-square method. For example, piano's fourth basic vector and cello's first basic vector, obtained with

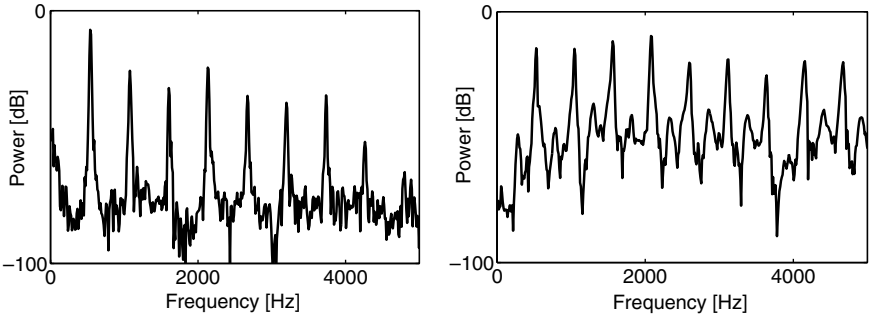
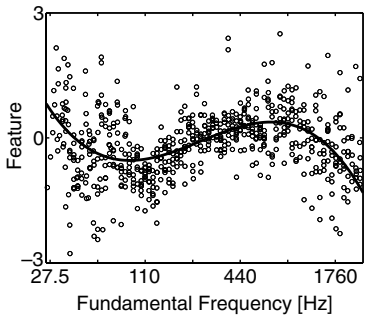
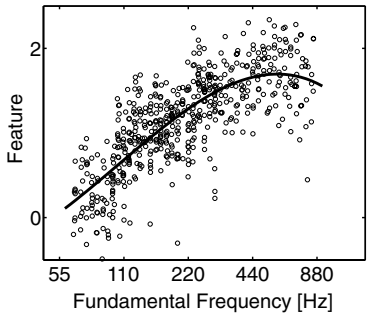



*Figure 1.* Power spectra of two different instruments at the onset time. These show that spectral peaks of an instrument with large non-harmonic components, such as (b) Alto Sax, has small kurtosis in contrast to those of an instrument with small non-harmonic components, such as (a) Recorder.




(a) Piano's 4th basic vector of features.




(b) Cello's 1st basic vector of features.

*Figure 2*. Examples of F0-dependent mean functions. Because the vertical axes of these graphs are not features defined in Section 2.1 but those obtained after standardization and dimensionality reduction (PCA and LDA), they do not have any units. In these graphs, we plot all the data of pianos or cellos in our database, while F0-dependent mean functions are estimated by only from training data in our experiments.

the two successive dimensionality reduction method (PCA and LDA), are depicted in Fig. 2(a) and (b), respectively. Then, the F0-normalized covariance is calculated by the following equation:

$$\Sigma_i = \frac{1}{n_i} \sum_{\boldsymbol{x} \in \chi_i} (\boldsymbol{x} - \boldsymbol{\mu}_i(f_{\boldsymbol{x}}))(\boldsymbol{x} - \boldsymbol{\mu}_i(f_{\boldsymbol{x}}))',$$

where $'$ is the transposition operator, $\chi_i$ and $n_i$ are the set of the training data of the instrument $\omega_i$ and its total number, respectively. $f_{\boldsymbol{x}}$ denotes the F0 of the feature vector $\boldsymbol{x}$. Because the F0-dependent mean function represents the pitch dependency of features, the F0-normalized covariance, obtained by subtracting the mean from each feature, eliminates the pitch dependency of features.

### *2.4. Application of the Bayes Decision Rule*

Once pitch and non-pitch dependencies of feature vectors are represented by the F0-dependent multivariate normal distribution, the Bayes decision rule is applied to identify the name or category of musical instruments. The discriminant function $g_i(\boldsymbol{x}; f)$ for the musical instrument $\omega_i$ is defined by

$$g_i(\boldsymbol{x}; f) = \log p(\boldsymbol{x} \mid \omega_i; f) + \log p(\omega_i; f), \quad (1)$$

where $\boldsymbol{x}$ is the feature vector of an input sound, $p(\boldsymbol{x} \mid \omega_i; f)$ is the probability density function (PDF) of this distribution and $p(\omega_i; f)$ is a priori probability of the instrument $\omega_i$.

The PDF of this distribution is defined by

$$p(\boldsymbol{x}|\omega_i; f) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2} D^2(\boldsymbol{x}, \boldsymbol{\mu}_i(f))\right\}, \quad (2)$$

where $d$ is the number of dimensions of the feature space and $D^2$ is the squared Mahalanobis distance defined by

$$D^2(\boldsymbol{x}, \boldsymbol{\mu}_i(f)) = (\boldsymbol{x} - \boldsymbol{\mu}_i(f))' \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i(f)).$$

Substituting Eq. (2) into Eq. (1), thus, generates the discriminant function $g_i(\boldsymbol{x}; f)$ as follows:

$$g_i(\boldsymbol{x}; f) = -\frac{1}{2} D^2(\boldsymbol{x}, \boldsymbol{\mu}_i(f)) - \frac{1}{2} \log |\Sigma_i| - \frac{d}{2} \log 2\pi + \log p(\omega_i; f).$$

The name of the instrument that maximizes this function, that is $\omega_k$ satisfying $k = \mathrm{argmax}_i\, g_i(\boldsymbol{x}; f)$, is determined as the result of musical instrument identification.

The a priori probability $p(\omega_i; f)$ represents whether the pitch range of the instrument $\omega_i$ includes $f$, that is,

$$p(\omega_i; f) = \begin{cases} 1/c & (\text{if } f \in R_i) \\ 0 & (\text{if } f \notin R_i) \end{cases}$$

where $R_i$ is the pitch range of the instrument $\omega_i$, and $c$ is the normalizing factor to satisfy $\sum_i p(\omega_i; f) = 1$.

## 3. Experiments and Results

### 3.1. Experimental Conditions

We conducted experiments on musical instrument identification for investigating improvement of the performance by the proposed method. We obtained the recognition rates by the commonly used multivariate normal distribution (called *baseline*) and by the proposed F0-dependent multivariate normal distribution, and compared them.

The benchmark used for evaluation is a subset of the "*RWC Music Database: Musical Instrument Sound*" (RWC-MDB-I-2001) [7], which is a large musical instrument sound databalse available to researchers around the world. This subset summarized in Table 1 was selected by the quality of recorded sounds and consists of 6,247 solo tones of 19 orchestral instruments. All data are sampled at 44.1 kHz with 16 bits. We first divided the whole data into 10 groups, and then repeated the following step 10 times: each time, we left out one of the 10 groups for training and used the omitted one for testing. That means that nine tenths of the data listed in Table 1 were used for calculating F0-dependent mean functions and F0-normalized covariances. This experiment technique is called 10-fold cross validation.

We evaluated the category-level performance of our method, because the category of instruments is useful for some applications including music retrieval. For example, when a user wants to find a piece of piano solo on a music retrieval system, the system can reject pieces containing instruments of different categories, which can be judged without identifying individual instrument names. We adopted the categories of musical instruments summarized in Table 2, which are determined based on the sounding mechanisms of instruments and existing studies [2, 5].

### 3.2. Results of Musical Instrument Identification

Table 3 summarizes recognition rates by both *baseline* and *proposed* methods. The proposed F0-dependent method improved the recognition rates at individual-instrument level from 75.73 to 79.73% and at category level from 88.20 to 90.65% in average. It also reduced recognition errors by 16.48 and 20.67% in average at individual-instrument and category levels, respectively.

*Table 1.* Contents of the database used in this paper.

| Instrument name (Abbrev.) | pitch range | # of tones | # of individuals | Intensity | Articulation |
|---|---|---|---|---|---|
| Piano (PF) | A0–C8 | 508 | 3 | Forte, normal | Normal only |
| Classical Guitar (CG) | E2–E5 | 696 | | & piano | |
| Ukulele (UK) | F3–A5 | 295 | | | |
| Acoustic Guitar (AG) | E2–E5 | 666 | | | |
| Violin (VN) | G3–E7 | 528 | | | |
| Viola (VL) | C3–F6 | 472 | | | |
| Cello (VC) | C2–F5 | 558 | | | |
| Trumpet (TR) | E3–A♯6 | 151 | 2 | | |
| Trombone (TB) | A♯1–F♯5 | 262 | 3 | | |
| Soprano Sax (SS) | G♯3–E6 | 169 | | | |
| Alto Sax (AS) | C♯3–A5 | 282 | | | |
| Tenor Sax (TS) | G♯2–E5 | 153 | | | |
| Baritone Sax (BS) | C2–A4 | 215 | | | |
| Oboe (OB) | A♯3–G6 | 151 | 2 | | |
| Faggoto (FG) | A♯1–D♯5 | 312 | 3 | | |
| Clarinet (CL) | D3–F6 | 263 | | | |
| Piccolo (PC) | D5–C8 | 245 | | | |
| Flute (FL) | C4–C7 | 134 | 2 | | |
| Recorder (RC) | C4–B6 | 160 | 3 | | |

*Table 2*. Categorization of 19 instruments.

| Category | Instruments (abbreviation) |
|---|---|
| Piano | Piano (PF) |
| Guitars | Classical Guitar (CG), Ukulele (UK), Acoustic Guitar (AG) |
| Strings | Violin (VN), Viola (VL), Cello (VC) |
| Brasses | Trumpet (TR), Trombone (TB) |
| Saxophones | Soprano Sax (SS), Alto Sax (AS), Tenor Sax (TS), Baritone Sax (BS) |
| Double Reeds | Oboe (OB), Faggoto (FG) |
| Clarinet | Clarinet (CL) |
| Air Reeds | Piccolo (PC), Flute (FL), Recorder (RC) |

The observation of these experimental results is summarized below:

- The recognition rates of six instruments (Piano (PF), Trumpet (TR), Trombone (TB), Soprano Sax (SS), Baritone Sax (BS), and Faggoto (FG)) were improved by more than 7%. In particular, the recognition rate for pianos was improved by 9.06%, and its recognition errors were reduced by 35.13%. This big improvement was attained since their pitch dependency is salient due to their wide range of pitch.
- The recognition rates for the four types of saxophones at individual-instrument level (47–73%) were lower than those at category level (77–92%). This is because sounds of these saxophones were quite similar. In fact, Martin reported that sounds of various saxophones are very difficult even for humans (music experts) to discriminate [5].
- Since we adopt the flat (non-hierarchical) categorization, the recognition rates at category level depend on the category. The recognition rates of GUITARS and STRINGS at category level were more than 94%, while those of BRASSES, SAXOPHONES, DOUBLE REEDS, CLARINET and AIR REEDS were about 70–90%. This is because instruments of these categories have similar sounding mechanism: these

*Table 3*. Accuracy by usual distribution (baseline) and F0-dependent distribution (proposed).

| | Individual-instrument level | | | Category level | | |
|---|---|---|---|---|---|---|
| | *Baseline* | *Proposed* | Improv. | *Baseline* | *Proposed* | Improv. |
| PF | 74.21% | 83.27% | +9.06% | 74.21% | 83.27% | +9.06% |
| CG | 90.23% | 90.23% | ±0.00% | 97.27% | 97.13% | −0.14% |
| UK | 97.97% | 97.97% | ±0.00% | 97.97% | 98.31% | +0.34% |
| AG | 81.23% | 83.93% | +2.70% | 94.89% | 95.65% | +0.76% |
| VN | 69.70% | 73.67% | +3.97% | 98.86% | 99.05% | +0.19% |
| VL | 73.94% | 76.27% | +2.33% | 93.22% | 94.92% | +1.70% |
| VC | 73.48% | 78.67% | +5.19% | 95.16% | 96.24% | +1.08% |
| TR | 73.51% | 82.12% | +8.61% | 76.82% | 85.43% | +8.61% |
| TB | 76.72% | 84.35% | +7.63% | 85.50% | 89.69% | +4.19% |
| SS | 56.80% | 65.89% | +9.09% | 73.96% | 80.47% | +6.51% |
| AS | 41.49% | 47.87% | +6.38% | 73.76% | 77.66% | +3.90% |
| TS | 64.71% | 66.01% | +1.30% | 90.20% | 92.16% | +1.96% |
| BS | 66.05% | 73.95% | +7.90% | 81.40% | 86.05% | +4.65% |
| OB | 71.52% | 72.19% | +0.67% | 75.50% | 74.83% | −0.67% |
| FG | 59.61% | 68.59% | +8.98% | 64.74% | 71.15% | +6.41% |
| CL | 90.69% | 92.07% | +1.38% | 90.69% | 92.07% | +1.38% |
| PC | 77.56% | 81.63% | +4.07% | 89.39% | 90.20% | +0.81% |
| FL | 81.34% | 85.07% | +3.73% | 82.09% | 85.82% | +3.73% |
| RC | 91.88% | 91.25% | −0.63% | 92.50% | 91.25% | −1.25% |
| Ave. | 75.73% | 79.73% | +4.00% | 88.20% | 90.65% | +2.45% |

*Baseline*: Usual (F0-independent) distribution.
*Proposed*: F0-dependent distribution.

categories are subcategories of "wind instruments" in conventional hierarchical categorization.
- The readers might think that a typical alternative approach of dealing with the pitch dependency could be a method that simply adds the value of the F0 to the feature vector (i.e., the number of dimensions of the feature vector is increased by one). We therefore compared this alternative method with our method. The experimental results showed that the alternative method is inferior to our method: the recognition rate was 75.88% on average, which is almost same with the baseline method.

## 4. Evaluation of the Bayes Decision Rule

The effect of the Bayes decision rule in musical instrument identification was evaluated by comparing with the $k$-NN rule ($k$-nearest neighbor rule; $k = 3$ in this paper) with/without LDA. Three variations of the dimensionality reduction are examined:

(a) Reduction to 79 dimension by PCA,
(b) reduction to 18 dimension by PCA, and
(c) reduction to 18 dimension by PCA and LDA.

The last one is adopted in the proposed method.

The experimental results listed in Table 4 showed that the proposed Bayes decision rule performed better in average than the 3-NN rule. Some observations are as follows:

- The Bayes decision rule with 79-dimension showed poor performance for Acoustic Guitar (AG), Trumpet (TR), Soprano Sax (SS), Tenor Sax (TS), Oboe (OB), and Flute (FL), since there are insufficient training data to estimate parameters of a 79-dimensional normal distribution. For small training sets with 79-dimension, $k$-NN is superior to the Bayes decision rule.
- LDA with the Bayes decision rule improved the accuracy of musical instrument identification from 66.50 to 79.73% on average. Although it seemed that PCA

*Table 4.* Accuracy by $k$-NN rule and the Bayes decision rule.

| | $k$-NN rule ($k = 3$) | | | Bayes decision rule | | |
|---|---|---|---|---|---|---|
| | 79-Dim. | 18-Dim. | | 79-Dim. | 18-Dim. | |
| | PCA | | PCA&LDA | PCA | | PCA&LDA |
| PF | 53.94% | 46.46% | 63.39% | 55.91% | 59.06% | 83.27% |
| CG | 79.74% | 77.16% | 75.72% | 98.28% | 97.27% | 90.23% |
| UK | 94.58% | 92.54% | 97.63% | 67.12% | 80.00% | 97.97% |
| AG | 95.05% | 92.79% | 97.00% | 19.97% | 44.14% | 83.93% |
| VN | 47.73% | 46.02% | 45.83% | 89.58% | 84.47% | 73.67% |
| VL | 55.93% | 54.24% | 61.86% | 71.19% | 79.24% | 76.27% |
| VC | 86.20% | 85.84% | 84.23% | 45.16% | 30.82% | 78.67% |
| TR | 36.42% | 38.41% | 47.02% | 41.72% | 72.85% | 82.12% |
| TB | 70.99% | 54.58% | 77.86% | 75.19% | 78.24% | 84.35% |
| SS | 23.08% | 14.20% | 24.85% | 48.52% | 66.86% | 65.89% |
| AS | 37.59% | 29.79% | 40.43% | 72.70% | 41.84% | 47.84% |
| TS | 62.09% | 66.01% | 68.63% | 30.07% | 61.44% | 66.01% |
| BS | 68.84% | 67.91% | 66.98% | 55.35% | 54.42% | 73.95% |
| OB | 47.68% | 48.34% | 49.01% | 43.71% | 81.46% | 72.19% |
| FG | 64.10% | 65.06% | 74.36% | 40.38% | 30.12% | 68.59% |
| CL | 93.45% | 87.93% | 93.10% | 95.51% | 93.45% | 92.07% |
| PC | 84.08% | 84.90% | 84.08% | 63.27% | 58.37% | 81.63% |
| FL | 88.06% | 72.39% | 94.03% | 35.82% | 84.33% | 85.07% |
| RC | 97.50% | 93.75% | 97.50% | 85.00% | 96.25% | 91.25% |
| Ave. | 70.27% | 66.98% | 72.53% | 62.11% | 66.50% | 79.73% |

with 79-dimension performed better than LDA for Classical Guitar (CG), Violin (VN), and Alto Sax (AS), the cumulative performance of LDA for the categories of strings and saxophones is better than that of PCA.

- We did not conduct the experiment using only LDA. This is because LDA cannot be applied to features that are highly correlative: the inverse matrix of the with-in covariance, which is used by LDA, is not accurately calculated when the feature space includes some highly correlative dimensions. Because PCA not only reduces the dimensionality but also orthogonalizes the feature space, for our features, some of which are highly correlative, using PCA before LDA is effective.

## 5. Conclusions

In this paper, we presented a method for musical instrument identification using the *F0-dependent multivariate normal distribution* which takes into consideration the pitch dependency of timbre. The method improved the recognition rates at individual-instrument level from 75.73 to 79.73%, and at category level from 88.20 to 90.65% on average, respectively. The Bayes decision rule with dimensionality reduction by PCA and LDA also performed better than the 3-NN method.

Future work will include to extend our method to deal with real-world music signals such as musical pieces and phrases containing several sounds of the same instrument. From the viewpoint of computational auditory scene analysis, it is also important to deal with a variety of sounds including non-musical sounds. We therefore plan to extend our method to deal with such general sounds because we can expect that our approach of handling the pitch dependency is also effective for general sounds that have pitches.

## Acknowledgments

We thank everyone who has contributed to building and distributing the RWC Music Database (Musical Instrument Sound: RWC-MDB-I-2001) [7]. We also thank Dr. Kazuhiro Nakadai and Dr. Hideki Asoh for their valuable comments.

## Note

1. In this paper, *pitch* and *timbre* are used as the meanings of fundamental frequency and acoustic features for discriminating instruments, respectively.

## References

1. J.C. Brown, "Computer identification of musical instruments using pattern recognition with cepstral coefficients as features," *Journal of Acoustic Society of America* vol. 103, no. 3, pp. 1933–1941, 1999.
2. A. Eronen and A. Klapuri, "Musical instrument recognition using cepstral coefficients and temporal features," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing, IEEE*, 2000, pp. 753–756.
3. I. Fujinaga and K. MacMillan, "Realtime recognition of orchestral instruments," in *Proceedings of International Computer Music Conference*, 2000, pp. 141–143.
4. K. Kashino, K. Nakadai, T. Kinoshita, and H. Tanaka, "Application of the bayesian probability network to music scene analysis," in *Computational Auditory Scene Analysis*, edited by D. Rosenthal and H. G. Okuno, Eds., Lawrence Erlbaum Associates, 1998, pp. 115–137.
5. K.D. Martin, "Sound-Source Recognition: A Theory and Computational Model," Ph.D. Thesis, MIT, 1999.
6. K. Kashino and H. Murase, "A sound source identification system for ensemble music based on template adaptation and music stream extraction," *Speech Communication*, vol. 27, nos. 3–4, pp. 337–349, 1999.
7. M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Music genre database and musical instrument sound database," in *Proceedings of International Conference on Music Information Retrieval*, 2003, pp. 229–230.
8. D. Rosenthal and H.G. Okuno, eds. *Computational Auditory Scene Analysis*, Lawrence Erlbaum Associates, Mahwah, New Jersey, 1998.

**Tetsuro Kitahara** received the B.S. from Tokyo University of Science in 2002 and the M.S. from Kyoto University in 2004. He is currently a Ph.D. course student at Graduate School of Informatics, Kyoto University. Since 2005, he has been a Research Fellow of the Japan Society for the Promotion of Science. His research interests include music informatics. He recieved IPSJ 65th National Convention Student Award in 2003, IPSJ 66th National Convention Student Award and TELECOM System Technology Award for

Student in 2004, and IPSJ 67th National Convention Best Paper Award for Young Researcher in 2005. He is a student member of IPSJ, IEICE, JSAI, ASJ, and JSMPC.

**Masataka Goto** received his Doctor of Engineering degree in Electronics, Information and Communication Engineering from Waseda University, Japan, in 1998. He then joined the Electrotechnical Laboratory (ETL; reorganized as the National Institute of Advanced Industrial Science and Technology (AIST) in 2001), where he has been engaged as a researcher ever since. He served concurrently as a researcher in Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Corporation (JST) from 2000 to 2003, and an associate professor of the Department of Intelligent Interaction Technologies, Graduate School of Systems and Information Engineering, University of Tsukuba since 2005. His research interests include music information processing and spoken language processing. Dr. Goto received seventeen awards including the IPSJ Best Paper Award and IPSJ Yamashita SIG Research Awards (MUS and SLP) from the Information Processing Society of Japan (IPSJ), Awaya Prize for Outstanding Presentation and Award for Outstanding Poster Presentation from the Acoustical Society of Japan (ASJ), Award for Best Presentation from the Japanese Society for Music Perception and Cognition (JSMPC), WISS 2000 Best Paper Award and Best Presentation Award, and Interaction 2003 Best Paper Award. He is a member of the IPSJ, ASJ, JSMPC, Institute of Electronics, Information and Communication Engineers (IEICE), and International Speech Communication Association (ISCA).

**Hiroshi G. Okuno** received the B.A. and Ph.D from the University of Tokyo in 1972 and 1996, respectively. He worked for Nippon Telegraph and Telephone, Kitano Symbiotic Systems Project, and Tokyo University of Science. He is currently a professor at the Department of Intelligence Technology and Science, Graduate School of Informatics, Kyoto University. He was a visiting scholar at Stanford University, and a visiting associate professor at the University of Tokyo. He has done research in programming languages, parallel processing, and reasoning mechanism in AI, and he is currently engaged in computational auditory scene analysis, music scene analysis and robot audition. He received the best paper awards from the Japanese Society for Artificial Intelligence and the International Society for Applied Intelligence, in 1991 and 2001, respectively. He edited with David Rosenthal "Computational Auditory Scene Analysis" from Lawrence Erlbaum Associates in 1998 and with Taiichi Yuasa "Advanced Lisp Technology" from Taylor and Francis Inc. in 2002. He is a member of IPSJ, JSAI, JSSST, JSCS, ACM, AAAI, ASA, and IEEE.