ARCHIVES OF ACOUSTICS
33, 1, 27–33 (2008)

AUTOMATIC IDENTIFICATION OF MUSIC PERFORMER USING THE LINEAR PREDICTION CEPSTRAL COEFFICIENTS METHOD

Magdalena CHUDY

Warsaw University of Technology
Institute of Radioelectronics
Nowowiejska 15/19, 00-665 Warszawa, Poland
e-mail: m.chudy@wsisiz.edu.pl

(*received June 15, 2007; accepted November 30, 2007*)

The paper describes a method of automatic identification of different music performers playing identical pieces of music on the same instrument. The performers' models based on the LPCC features and vector quantization are proposed as methods of classification. The presented approach was verified with a database of experimental samples of Bach's *1st Cello Suite* recorded especially for this study and the original audio CD recordings of Bach's *6 Cello Suites* performed by six famous cellists.

Keywords: linear prediction cepstral coefficients, music performer identification, vector quantization.

1. Introduction

In the time of a rapid progress of the computer and Internet technology and with easy access to a large amount of music data on the Internet, the music performer identification becomes a more and more popular task in the domain of music information retrieval. Techniques for automatic recognition of pop singers or artists as well as for classical music performers are strongly needed in order to effortlessly document unlabeled or inexactly labeled data [3, 5, 6, 9, 10].

In terms of the goal, a singer recognition is analogous to a speaker recognition, and by the assumptions that creating sound on a string instrument is similar to the process of speaking and that every musician plays the instrument with a unique timbre of sound, an instrumentalist recognition is analogous to a speaker recognition too. Success in solving these problems depends on the detection and exploitation of the characteristic features that distinguish one person's "sound" or voice from another's. Preliminary results of exploring the problem of automatic identification of a music performer were described in author's previous contributions to this subject [1, 2].

In a music performer identification system, a feature extraction module obtains a performer's features from audio signals. These features capture acoustic characteristics of the performer's sound timbre. A method based on LPCC (*Linear Prediction based Cepstrum Coefficients*) is used. In a speaker recognition task, a total number of 18–20 LPCCs computed from each speech signal frame seemed to be enough to acquire voice spectral information and to distinguish speakers with high accuracy. To efficiently model a music signal, a total number of 71 LPCCs is assumed [1].

2. Music performer characteristics modeling

With LPCC-based feature vectors extracted from audio frames, the vector quantization approach is applied to build the acoustic model of a music performer. This method approximates the training data with a so-called vector quantizer codebook (or *dictionary*) by minimizing a MSE criterion. The codebook consists of a certain number of code vectors (L value), so-called "codewords". For one of the feature vectors sequences used as a training set and initial codebook, an iterative process is performed, following the Generalized Lloyd Algorithm (GLA) steps, to design a final music performer's codebook. For each iteration of the GLA, each training vector is associated with its nearest codeword by calculating the distance. Euclidean squared error is the distance measure. Once all the training vectors has been associated with their nearest codeword vectors, the mean squared error for the codebook is calculated (quantization error) and it is checked to see if the stopping criteria for the process has been satisfied. In finally calculated codebook every codeword vector is associated with k training vectors [4].

The effective identification process depends on the codebook size parameter. For speaker recognition, the most efficient L value ranges from 8 to 32 vectors. In the case of instrumentalist's identification, L value of 64 vectors was empirically obtained [1, 2].

3. Music performer identification procedure

Music performer identification refers to the task of determining who among a group of candidate performers has played a given part of a musical piece. This involves an N-class decision, where N is the number of candidate performers. For every musician in a music performer identification system, his optimal 64-vector codebook is designed. In fact, the identification process corresponds to the test sequence quantization process. It depends on finding (for every test feature vector in the test sequence) the nearest codeword vectors in the performers' codebooks by calculating Euclidean squared error as Euclidean distance measure between them. The system identifies the performer for whom a normalized summary distance (quantization error) is minimal in the whole test sequence.

4. Experimental results

4.1. Music database

The music data used in this study consisted of 222 clips extracted from experimental studio recordings of two parts of Bach's *1st Cello Suite* made by six performers on two different cellos with every part of a piece played in two versions as well as the original CD recordings of Bach's *6 Cello Suites* performed by six famous cellists. They were grouped into three data sets and labeled as *1st cello*, *2nd cello* and CD recordings sets. In *1st cello* and *2nd cello* recordings sets, only two sets consisted of 9 and 7 clips, all the others included 14 clips. In the CD recordings set the performers data sets consisted of 9 or 12 clips. The length of the clips ranged from 5 to 10 seconds.

All the clips were captured in a mono channel *.wav* file format at a sampling rate of 44.1 kHz and 16 bits per sample. Feature vectors, each consisting of 71 LPCCs for every signal frame, were extracted from this data using a 1024-sample Hamming-windowed frames with 512-sample shifts. Consequently a 10-second clip corresponded to 861 feature vectors sequence. For the whole music database the length of the sequence fluctuated respectively from 410 to 861 feature vectors.

Table 1. Artists and recordings.

Artist Index	Artist	CD Recording
AB	Andrzej Bauer	CD ACCORD, Bach *Cello Suites*, ACD 032, 1999
MM	Mischa Maisky	DG, Bach *6 Cello Suiten*, 463 314-2, 2000
BM	Barbara Marcinkowska	SEPMQUANTUM, Bach *Cello Suites Nos.1-3*, DQM 6972, 1996
IM	Ivan Monighetti	DUX Recording Producers, Bach *Six Cello Solo Suites*, DUX 0301/0302, 2002
PT	Paul Tortelier	EMI Classics, Bach *Cello Suites Nos. 1-3*, 5 73526 2, 1999
PW	Pieter Wispelwey	CHANNEL Classics, Bach *6 Suites per violoncello solo*, CCS 12298, 1998

4.2. Music performer identification results

Three identification experiments were performed on the data sets. The first two experiments aimed to investigate the effectiveness of the proposed method on each cello set separately. The third series of tests dealt with the famous artists' recognition and with studying the so-called *album effect* on the CD recordings set. The experiments were conducted in a double leave-one-out manner. At first, each of the six performers became a target (one at a time), and the procedure rotated through all of them. Then, each of the considered performer recording clips was used for building the performer's acoustic model, and the remaining clips were used as a test set. The procedure was applied to all of the performer's clips. The identification accuracy was computed as the percentage of correctly-identified clips over the total number of test clips. Tables 3–5 show the confusion matrices obtained for *1st cello*, *2nd cello* and the CD recordings set respectively.

Table 2. Database description.

Music piece title	*1st cello* and *2nd cello* recordings sets clips' labels	CD recordings set clips' labels
Suite No. 1 in G – Prélude	Suita_01_Prel1 bars 1–4	Suita_01_Prel1 bars 1–4
	Suita_01_Prel2 bars 5–7	Suita_01_Prel2 bars 5–7
	Suita_01_Prel3 bars 8–10	Suita_01_Prel3 bars 8–10
	Suita_01_Prel4 bars 11–13	Suita_01_Prel4 bars 11–13
	Suita_01_Prel5 bars 14–17	Suita_01_Prel5 bars 14–17
Suite No. 1 in G – Gigue	Suita_01_Gigue1 bars 1–4	Suita_01_Gigue1 bars 1–4
	Suita_01_Gigue2 bars 5–8	Suita_01_Gigue2 bars 5–8
Suite No. 2 in d – Prélude		Suita_02_Prel1 bars 1–4
Suite No. 3 in C – Prélude		Suita_03_Prel1 bars 1–6
Suite No. 4 in Es – Prélude		Suita_04_Prel1 bars 1–8
Suite No. 5 in c – Prélude		Suita_05_Prel1 bars 1–3
Suite No. 6 in D – Prélude		Suita_06_Prel1 bars 1–4

From the results presented in Tables 3 and 4, one can conclude that the effective identification of music performers playing the same instrument does not depend on the instrument but evidently varies from performer to performer. The total identification accuracy obtained for both instruments is very similar (about 88%). It was possible to achieve these comparable results on condition that the same acoustic environment was provided for all of the recorded players. On the other hand, each of the original CD recordings was arranged in a completely different acoustic space, every artist played his own unique instrument, and different mastering and post-processing techniques were applied. This classification factor named the *album effect* can affect identification performance [9]. Consequently, the classifier identifies the album instead of the instrumentalist. The album effect as well as the performer's manner of playing causes that the performers sound characteristics strongly differ from each other and the total identification accuracy (Table 5) is much higher (about 94%). In two cases it reaches even 100% matches.

When applying a vector quantization approach to classification tasks, the influence of FVs sequences length on the identification performance was observed, especially in a vector quantizer codebook designing phase. As it appeared, the capability of LPCCs

Table 3. Confusion matrix of the performer identification tested on *1st cello* recordings set.

Performer of the test clip		Results of the performer identification						Accuracy (%)
Performer Index	Total number of performer test clips	Performer Index						
		1	2	3	4	5	6	
1	182	**167**	3	4	1	4	3	91.8
2	182	8	**169**			5		92.9
3	72	3	3	**59**		1	6	81.9
4	182	3	33	1	**137**	2	6	75.3
5	182	3	9	2	2	**166**		91.2
6	182	4	1				**177**	97.3
Total number of test clips	**982**	Total identification accuracy (%)						**88.4**

Table 4. Confusion matrix of the performer identification tested on *2nd cello* recordings set.

Performer of the test clip		Results of the performer identification						Accuracy (%)
Performer Index	Total number of performer test clips	Performer Index						
		1	2	3	4	5	6	
1	182	**179**		1	1		1	98.4
2	42		**38**		3		1	90.5
3	182	1	10	**154**	4	4	9	84.6
4	182	6	6	11	**154**	2	3	84.6
5	182	18	1	2	1	**156**	4	85.7
6	182	4	9	7	2	6	**154**	84.6
Total number of test clips	**952**	Total identification accuracy (%)						**88.1**

vectors sequence to capture the performer timbre characteristics from each tested music clip was prior to its length parameter. The total performer identification accuracy varied from one testing routine to another and the length of a training sequence for a vector quantizer codebook designed as a performer model seemed not to be significant. The performed experiments proved that the expressive language aspects of music such as tempo, dynamics, articulation, and the musical structure elements such as rhythm and tonality are not significant factors for the identification effectiveness either.

Looking at Tables 3 and 4, one can notice a tendency to misidentify one performer as another. The suggestion is, that the lower identification accuracy might be attributed to the presence of the higher variations of timbre characteristics in the performer's training sequence. The performers associated with a lower identification accuracy had usually a higher quantization error in a vector quantizer codebook designing phase.

Finally, it is important to note, that the performer's "sound" characteristics captured on one instrument differ from those captured on another. To solve the performer classification problem in $N \times M$ class, where M is the number of allowed instruments, the inclusion of music clips from both instruments recordings sets is required in the performer's codebook designing phase.

Table 5. Confusion matrix of the performer identification tested on CD recordings set.

Performer of the test clip		Results of the performer identification						Accuracy (%)
Artist Index	Total number of artist test clips	Artist Index						
		AB	MM	BM	IM	PT	PW	
AB	132	**131**				1		99.2
MM	132	6	**105**	4	3	14		79.5
BM	72		2	**70**				97.2
IM	132		14		**117**		1	88.6
PT	72					**72**		100.0
PW	132						**132**	100.0
Total number of test clips	**672**	Total identification accuracy (%)						**94.1**

5. Conclusions and future works

In this paper, a method based on LPCC features for the purpose of music performer automatic identification was introduced and verified. In relation to the former study [2], the music database was extended by adding another version of each musical piece to *1st* and *2nd cello* recordings sets. Secondly, two full-length testing procedures were performed on both sets to compare the identification effectiveness depending on a played instrument. The third experiment was performed to observe the influence of the so-called album effect on the classification task using CD recordings set.

It was shown, that classical music performers in solo music recordings can be distinguished from each other by their "sound" characteristics. The instrumentalist "sound" characteristics can be extracted from music signals and efficiently modeled by traditional spectral features like LPCCs. The achieved identification accuracy is about 88% for experimental studio recordings and about 94% for original CD albums. The results are consistent with those obtained and described in author's earlier publications [1, 2]. For the topic of the classification task, this study confirmed the advantages of the vector quantization approach which simplifies the identification procedures and does not require additional computing complexity.

Although the proposed solutions led to successful results, they may only be treated as preliminary investigation in the task of a music performer identification. The fundamental problem due to strong correlation between the performer and the instrument is clearly visible and should be solved. Future work needs to concentrate on the separation of the human and instrument factors from the music sound spectral characteristics. Obviously, to obtain significant classification results more advanced classification methods are required.

References

[1] CHUDY M., *System for automatic identification of music performer* (*System do automatycznego rozpoznawania wykonawcy utworu muzycznego*), M.S. Degree thesis, Computer Science Department, Warsaw School of Information Technology, Warsaw 2006.

[2] CHUDY M., *Automatic identification of music performer* (*Automatyczna identyfikacja wykonawcy utworu muzycznego*), Proceedings of the 11th Symposium AES "New Trends in Audio and Video", p. 170–175, Bialystok 2006.

[3] FUJIHARA H., KITAHARA T., GOTO M. *et al.*, *Singer identification based on accompaniment sound reduction and reliable frame selection*, Proceedings of the 6th International Conference on Music Information Retrieval, London 2005.

[4] GERSHO A., GRAY R., *Vector quantization and signal compression*, Kluwer Academic Publishers, Boston 1992.

[5] KIM H. G., BERDAHL E., MOREAU N., SIKORA T., *Speaker recognition using MPEG-7 descriptors*, Eurospeech-2003, pp. 489–492, Sep. 2003.

[6] KIM Y. E., WHITMAN B., *Singer identification in popular music recordings using voice coding features*, Proceedings of the 3rd International Conference on Music Information Retrieval, Paris 2002.

[7] MCKINNEY M. F., BREEBAART J., *Features for audio and music classification*, Proceedings of the 4th International Conference on Music Information Retrieval, Baltimore, Maryland (USA) 2003.

[8] MESAROS A., ASTOLA J., *The mel-frequency cepstral coefficients in the context of singer identification*, Proceedings of the 6th International Conference on Music Information Retrieval, London 2005.

[9] NWE T. L., LI H., *Exploring vibrato-motivated acoustic features for singer recognition*, IEEE Transactions on Audio, Speech, and Language Processing, **15**, 2, 519–530 (2007).

[10] TSAI W. H., WANG H. M., *Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals*, IEEE Transactions on Audio, Speech, and Language Processing, **14**, 1, 330–341 (2006).