Group Delay Function from All-Pole Models for Musical Instrument Recognition

Aleksandr Diment[1], Padmanabhan Rajan[2],
Toni Heittola[1], and Tuomas Virtanen[1]*

[1] Department of Signal Processing,
Tampere University of Technology, Finland
[aleksandr.diment,toni.heittola,tuomas.virtanen]@tut.fi
[2] School of Computing and Electrical Engineering,
Indian Institute of Technology Mandi, Himachal Pradesh, India
padman@iitmandi.ac.in

Abstract. In this work, the feature based on the group delay function from all-pole models (APGD) is proposed for pitched musical instrument recognition. Conventionally, the spectrum-related features take into account merely the magnitude information, whereas the phase is often overlooked due to the complications related to its interpretation. However, there is often additional information concealed in the phase, which could be beneficial for recognition. The APGD is an elegant approach to inferring phase information, which lacks of the issues related to interpreting the phase and does not require extensive parameter adjustment. Having shown applicability for speech-related problems, it is now explored in terms of instrument recognition. The evaluation is performed with various instrument sets and shows noteworthy absolute accuracy gains of up to 7% compared to the baseline mel-frequency cepstral coefficients (MFCCs) case. Combined with the MFCCs and with feature selection, APGD demonstrates superiority over the baseline with all the evaluated sets.

Keywords: Musical instrument recognition, music information retrieval, all-pole group delay feature, phase spectrum

1 Introduction

Musical instrument recognition is one example of the subtopics of music information retrieval, and it has been most actively explored since the 1990's when the systems aimed at handling small numbers of instruments represented by isolated notes were already reaching impressive performance scores of 98% [17] and 100% [16]. During the following years, various systems realising numerous methods and applied for different numbers of instruments have been developed.

* This research has been funded by the Academy of Finland, project numbers 258708, 253120 and 265024.

Published in the CMMR 2013 post proceedings, LNCS series, Springer Verlag, 2014. The final publication is available at link.springer.com.

The recent works on the subject utilise existing classification methods, which are novel in terms of the given problem (e.g., semi-supervised learning [5] and applying missing feature approach for polyphonic instrument recognition [11]), as well as introduce new features (e.g., multiscale MFCCs [26] and amplitude envelopes of wavelet coefficients [13]). There exists an established set of features commonly applied for instrument recognition. Depending on whether they treat audio from the temporal or spectral point of view, these are subcategorised accordingly.

The temporal features address instrument recognition under the assumption that the relevant information is within the transient properties of a signal. Such assumption is perceptually motivated: the attack characteristics are believed to play crucial role in human recognition of musical instruments [7].

The spectral features employ a different approach. Particularly, those that are related to the harmonic properties of a sound (e.g., inharmonicity and harmonic energy skewness) do preserve the important properties of the musical instrument timbre [1]. The same applies to other spectrum-related features as well, such as mel-frequency cepstral coefficients (MFCCs). However, being spectrum-based, these features tend to concentrate only on the magnitude part of the spectrum.

Spectral information is complete only if both magnitude and phase spectra are specified. Signal processing difficulties, such as wrapping of the phase and dependency of the phase on the window position, make direct processing of the phase spectra challenging. A popular solution is to use the *group delay function*, which is defined as the negative derivative of the phase spectrum. The group delay function is well-behaved only if the zeros of the system transfer function are not close to the unit circle.

One example of overcoming the latter difficulty is by computing the so-called *modified group delay function* [24], which has been applied for speech recognition [15, 24], as well as recently for musical instrument recognition [6]. It has shown in the case of instrument recognition a comparable performance with such established features as MFCCs and an up to 5.1% recognition accuracy improvement in combination with MFCCs. However, the parameters of the function need to be adjusted to the specific application scenario, which is computationally expensive, so a simpler approach is desirable.

This work proposes utilising a more elegant method of acquiring a well-behaved group delay feature, which is novel in instrument recognition and does not require as much adjustment of parameters to an application scenario. The main aspect of this method is to calculate the group delay function from all-pole models of a signal, formed by linear predictive analysis (later referred to as APGD, all-pole group delay). While achieving the same goal of overcoming the irrelevant high amplitude spurious peaks in the group delay function, it appears more universally applicable and, therefore, beneficial. Previously, this method has shown to be successful in formant extraction [27] and speaker recognition [25].

The calculation of the APGD feature is proposed for pitched instrument recognition, either primarily or as a complement to the established MFCCs, motivated by the fact that additional information relevant in terms of musical instrument classification is concealed in phase. Its performance is evaluated in separate note classification scenarios with instrument sets of different sizes.

The paper is organised according to the following structure. Section 2 presents the motivation for computing group delay as a feature in general and the APGD version in particular, as well as its calculation procedure. Subsequently, Section 3 introduces the implemented instrument recognition system, which incorporates APGD calculation as its feature extraction block. Its performance is consecutively evaluated in Section 4. Finally, the conclusions about the applicability of the feature are drawn along with the future research suggestions in Section 5.

2 Group Delay Function

In this section, firstly, a motivation for utilising phase information for musical instrument recognition is stated along with the reasoning for computing APGD in particular. Subsequently, the details of calculation of the group delay function and its APGD extension are presented.

2.1 Motivation for Musical Instrument Recognition

Phase is often overlooked in many audio processing solutions due to the complications related to the unwrapping of the phase spectrum. In spite of that, phase could be highly informative due to its high resolution and ability of indicating peaks in the magnitude spectrum envelope. In terms of speech-related problems, these correspond to formants, useful for extracting speech content. There has been studies [2] showing that phase contributes significantly to speech intelligibility, contrary to the common notion of its perceptual negligibility.

In the musical instrument signals, however, the presence of formants in the spectrum is not as strong [16], or they are not a factor independent from fundamental frequency, in contrast to speech signals. For example, in the spectra of trombone or clarinet, due to the acoustical change of active volume of their body during the sound production, the resonances depend on pitch [9, 20].

Nevertheless, a phase-based feature is applicable for instrument recognition as well. Broadly speaking, while the commonly applied MFCCs feature is capable of modelling the resonances introduced by the filter of the instrument body, it neglects the spectral characteristics of the vibrating source, which also play their role in human perception of musical sounds [10]. Incorporating phase information attempts to preserve this neglected component.

Furthermore, considering instruments with such resonators as stretched strings and air columns in pipes, their natural resonances are not perfectly harmonic. However, due to such phenomenon as *mode locking*, individual modes of such instruments are locked into the precise frequency and phase relationships, leading to repeating waveforms of sustained tones of these instruments. This phenomenon occurs in case certain conditions favouring the effect are met [9]. A phase-related feature could aid in capturing the presence of this effect.

2.2 Group Delay Function

The *group delay function* is of a signal $x[n]$ obtained as [3]

$$\tau_g(\omega) = -\mathrm{Im}\left(\frac{d}{d\omega}\log(X(\omega))\right) \tag{1}$$

$$= \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{|X(\omega)|^2}, \tag{2}$$

where $X(\omega)$ and $Y(\omega)$ are the Fourier transforms of $x[n]$ and $y[n]$, and $y[n] = nx[n]$. The advantage of Equation (2) over the conventional way to obtain phase information is that no explicit unwrapping is needed.

The group delay function is well-behaved only if the zeros of the system transfer function are not close to the unit circle. The zeros may be introduced by the excitation source or as a result of short time processing [4,15]. When zeros of the transfer function are close to the unit circle, the magnitude spectrum exhibits dips at the corresponding frequency bins. Due to this, the denominator term in Equation (2) tends to a small value, resulting in a large value of the group delay function $\tau_g(\omega)$. This manifests itself in spurious high amplitude spikes at these frequencies, masking out the resonance structure in the group delay function.

One way of addressing this issue is by introducing a modification [24] of the group delay function (MODGDF), which suppresses the zeros of the transfer function. This is done by replacing the magnitude spectrum $X(\omega)$ by its cepstrally smoothed version $S(\omega)$. Two additional parameters are introduced to control the dynamic range.

Although this has shown to be a reasonable approach, applicable among others for musical instrument recognition [6], the presence of the three parameters that need to be adjusted to an environment does not necessary appear desirable due to the computational requirements such parameter tuning imposes. Another way of obtaining a group delay function, which lacks of this complication, is the group delay function of all-pole models.

2.3 Group Delay Function of All-Pole Models

By modelling a musical instrument with a a source-filter model [19] and assuming the filter all-pole, the spectrum of the such filter may be approximated with aid of linear prediction. The latter has been shown to be an efficient tool for the analysis of sounds of musical instruments, whose transient part is significant in terms of tone quality, such as piano [23]. Another example where all-pole modelling has been used for analysis of musical instrument sounds is the modelling of the guitar body response [18].

Linear prediction is formulated as [21]

$$H(\omega) = \frac{G}{1 - \sum_{k=1}^{p} a(k)e^{-j\omega k}}. \tag{3}$$

The coefficients $a(k)$ are determined by the method of least squares in such a way that the power spectrum of $H(\omega)$ matches the power spectrum of the signal

$|X(\omega)|^2$. The all-pole group delay function is computed from the phase response of this filter formed by $H(\omega)$.

Figure 1 shows the magnitude spectra and corresponding all-pole group delay of one frame of a one note produced by piano. One of the fundamental properties of the group delay function is its high resolution, which makes the formants visible and contributes additional information to the magnitude spectrum. Indeed, in the figure, one may observe the clearly emphasised formants in the APGD plot, not as easily seen in magnitude spectrum, which makes the function helpful for an instrument classification problem.

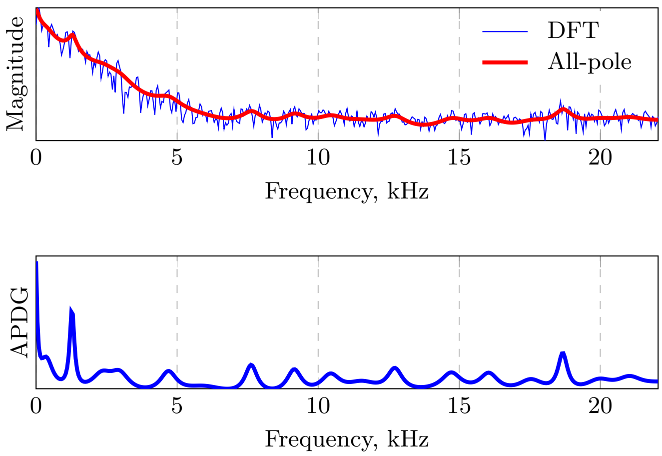

Fig. 1. A frame of a piano note “A0” (MIDI note 21) played in normal playing style with dynamics forte: its magnitude spectra (DFT and all-pole, upper panel) and a group delay function, model order 40 (lower panel).

To convert the all-pole group delay function into a feature, a discrete cosine transform (DCT) is applied. This performs decorrelation, and a certain number of coefficients are retained, excluding the zeroth. The feature is calculated in short frames under the assumption of spectral stationariness within their length, and the Fourier analysis is performed with the aid of DFT. The overal calculation procedure of APGD, illustrated by a block-diagram in Figure 2, is the following.

1. Perform all-pole modeling on the frame. Obtain the filter coefficients $a(k)$.
2. From the $a(k)$, form the frequency response $H(\omega)$ using Equation (3) with $G = 1$ (for a simplified representation capturing formant locations).
3. Compute the group delay function by taking the negative derivative of the phase response of $H(\omega)$. In practice, the derivative is computed using the sample-wise difference.
4. Take DCT on the group delay function and keep a certain number of coefficients, excluding the zeroth.
5. Delta coefficients are appended to the feature in a conventional manner.

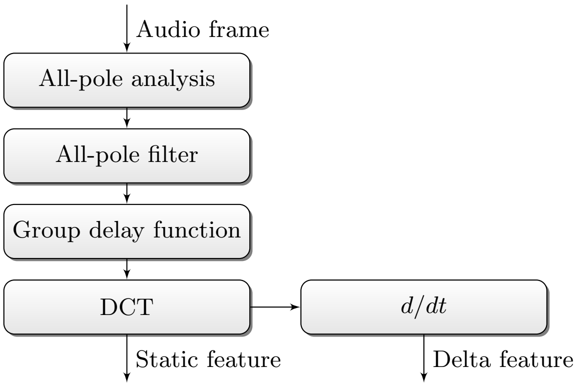

Fig. 2. A block diagram of the calculation of APGD.

3 System Description

The details of the developed musical instrument recognition system that incorporates APGD as one of its features are addressed in this section. The simplified block diagram of the system is presented in Figure 3, and the upcoming subsections are following the implementation of its building blocks.

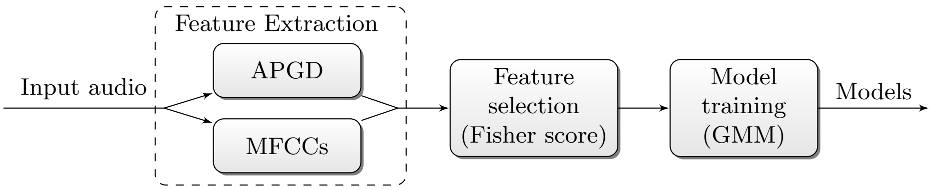

Fig. 3. A block diagram of the feature extraction and training phases.

3.1 Feature Extraction

As primarily explored features, APGD, as well as its first derivative, are incorporated in the calculation. Additionally, a baseline scenario is included, i.e., the calculation of the static and delta MFCCs. Those are currently quite commonly applied for musical instrument recognition, proven to be amongst the most effective features [8] due to their ability to parametrise the rough shape of the spectrum, which is different for each instrument. The mel transformation, which is included in the calculation of MFCCs, is based on human perception experiments and has been demonstrated to effectively represent perceptually important information in terms of instrument recognition [22]. The classification results produced by recognisers based on MFCCs have been shown to resemble human classifications in terms of the similar confusions [8]. This baseline MFCCs scenario is intended to indicate the expected performance of the system with the given data when utilising such established feature.

Frame-blocking is performed with 50%-overlapping windows of length 20 ms. The number of mel filters in MFCCs calculation is 40, and the number of extracted coefficients is set to 16 with exclusion of the zeroth (i.e., resulting in 15 static and 15 delta coefficients). In the case of APGD, 60 static and 60 delta coefficients are extracted as suggested by preliminary experiments. For the value of the LPC order, a parameter search has been performed within the range 20 to 70, and the order of 40 has been selected.

The calculation of the combination of these features is foreseen in order to investigate whether the APGD, if not as effective as MFCCs *per se*, is capable of enhancing the performance of the system when used as a complement to the baseline feature. This way, a feature set that incorporates both amplitude and phase information is acquired. The combined features of dimension 150 are obtained by concatenating the values of MFCCs and APGD.

3.2 Feature Selection

Generally, feature selection is applied in order to keep those features that are more relevant for the separability of classes. In this study, one of the goals is to demonstrate the effectiveness of the combined method when the dimensionality is reduced down to the size of the baseline MFCCs feature vectors.

The chosen approach for feature selection is the Fisher score [14]:

$$F_r = \frac{\sum_{i=1}^{c} n_i(\mu_i - \mu)^2}{\sum_{i=1}^{c} n_i \sigma_i^2}, \tag{4}$$

where r is the index of the feature being scored, $i = 1, \ldots, c$ is the class index, μ_i and σ_i^2 are the mean and the variance of class i, and μ is the mean of the whole set. The highest score is assigned to the feature on which the data points of different classes are far from each other, and the points of the same class are close to each other. A fixed number of features having the highest score are selected.

3.3 Training and Recognition

The training and recognition phases are performed by employing Gaussian mixture models (GMM). For each class, the feature vectors from the training data are used to train the GMM, i.e., to estimate the parameters of such model that best explains these features. The expectation-maximisation (EM) algorithm is used for this purpose, and each class is represented by a GMM of 16 components.

In recognition, the trained models of each class are fit into each frame of the test instances, producing log-likelihoods. Log-likelihoods are accumulated over the frames of the test instance. Thereupon, the label of the class whose model has produced the highest log-likelihood is assigned to that instance.

4 Evaluation

The performance of the proposed approach is evaluated in a separate note-wise instrument classification scenario. Several instrument sets grouped by the level of complexity of the resulting problem are considered. The instrument content of these sets is presented below, followed by the obtained evaluation results.

4.1 Acoustic Material

The recordings (sampling frequency 44.1 kHz) used in evaluation originate from the RWC Music Database [12]. Each of the instruments is represented in most cases by three instances, which stand for different instrument manufacturers and musicians. These are subdivided into subsets according to the playing styles (e.g., bowed vs plucked strings), and only one playing style per instrument is taken into account. In total, three instruments sets (Table 1) are considered, consisting of 4, 9 and 22 instruments. The choice of instruments in the first two sets is influenced by the requirement of a sufficiently high number of notes per instrument for its consistent representation. The largest set, composed of diverse instruments and even vocals, not necessary sufficiently represented in the database, is intended to demonstrate a highly complex classification scenario.

Table 1. Instrument sets used in evaluation.

Set	List of instruments
4-set	Acoustic Guitar, Electric Guitar, Tuba, Bassoon
9-set	Piano, Acoustic Guitar, Electric Guitar, Electric Bass, Trombone, Tuba, Bassoon, Clarinet, Banjo.
22-set	"4-set" + "9-set" + "woodwinds" (Oboe, Clarinet, Piccolo, Flute, Recorder) + "strings" (Violin, Viola, Cello, Contrabass) + vocals (Soprano, Alto, Tenor, Baritone, Bass)

The dataset, where each instrument is represented by several hundred recordings, is randomly divided into the training and test subsets. The subsets are acquired from different instrument instances in order to resemble a real-life application scenario. The ratio between the sizes of the training and test subsets is roughly 70%/30%.

4.2 Results

The evaluation results obtained with each of the instrument sets are summarised in Table 2. The values of the accuracy of MFCCs are somewhat different (within the range of 1%) from the previously reported [6], although the same database and the same instruments were used. This is due to the fact that the corresponding tests needed to be repeated in order to report additional, more detailed results. The randomisation of the separate note recordings that occurs during the

Table 2. Evaluation results, where the performance of APGD is compared to the performance of MODGDF [6]. FeatSel 30 and 120 stand for applying the feature selection and selecting 30 and 120 features, respectively.

Method	**Recognition accuracy, %**		
	4-set	9-set	22-set
MFCCs	90.9	83.7	68.8
MODGDF	84.4	59.9	41.7
APGD	97.9	84.8	63.3
MODGDF + MFCCs	96.0	84.9	70.7
APGD + MFCCs	93.7	87.0	68.3
APGD + MFCCs + FeatSel 30	95.5	84.2	66.2
APGD + MFCCs + FeatSel 120	**94.4**	**85.7**	**70.0**

division of the dataset into training and test sets, as well as randomisation during the initialisation of the EM algorithm are the reasons for such behaviour. The overall trend of accuracy improvements along the evaluated scenarios, however, is the same, and the differences between the performance of presented methods are of the same character.

The results include among others applying feature selection methodology with selecting 30 and 120 features. In the first case, the idea is to demonstrate the effectiveness of the combined approach when the data dimensionality is reduced down to the size of MFCCs vectors. In the case of 120 selected features, shown to be an optimal parameter during preliminary tests, the goal is to maximise accuracy gains.

By examining the obtained results, one may observe that APGD, used as such, is capable of serving as a reliable feature. The improvement over the baseline MFCCs scenario is apparent with both 4- and 9-instrument sets. The more real-life application case, namely, the set of 22 instruments, has shown to be somewhat complicated for the APGD feature, showing a decrease in accuracy of 5.5%. However, by combining both features and applying feature selection, the accuracy improvement compared to the baseline case is, nevertheless, achieved. The combined method with feature selection has demonstrated to be effective even when the dimensionality of the feature space is reduced down to the size of the baseline MFCCs vectors, i.e., when no more degrees of freedom is introduced. This shows that APGD does indeed provide new information in terms of musical instrument classification, which is explained both by the relevance of LPC analysis of musical instrument sounds and its capability of capturing phase information, which the conventional features tend to neglect.

A somewhat more specific comparison of the features can be performed by observing the instrument-wise accuracies. As seen in Figure 4, obtained with the set of nine instruments, the accuracy improvement introduced by APGD-based methods compared to MFCCs is mostly pronounced in the cases of string instruments (Acoustic Guitar, Electric Guitar, Electric Bass and Banjo), which

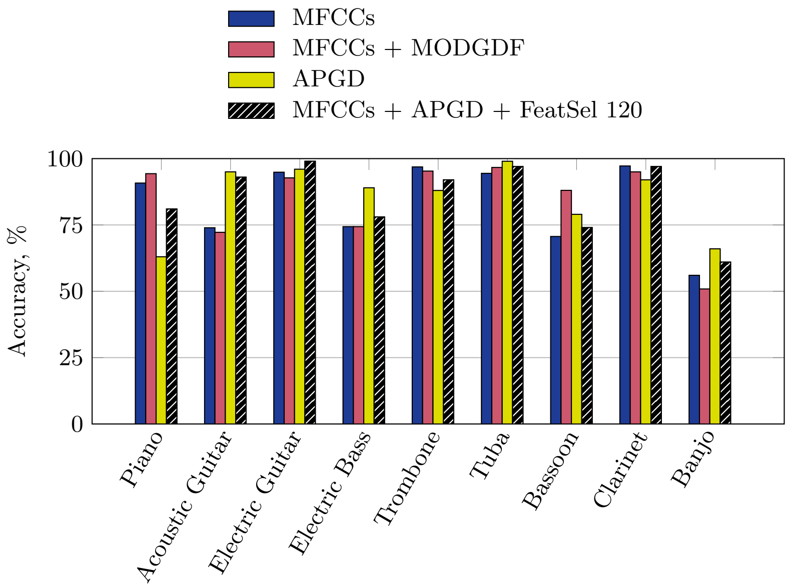

Fig. 4. Instrument-wise accuracies in selected evaluation scenarios with the 9 instruments set.

is in agreement with the reported applicability of all-pole modelling of guitar body response [18]. Additionally, some improvement is noticeable in the cases of instruments Tuba and Bassoon, which may be interpreted by phase-based nature of APGD, since the improvement in these instruments is also apparent in the case of MODGDF.

Even more thorough analysis of the results may be performed by addressing the confusion matrices, presented in Table 3. Amongst the most obvious observations one may mention how the APGD-based method missclassifies the instrument Piano for Banjo, whereas MFCCs tend to classify Piano correctly in most of the cases. However, by applying a combined method with feature selection, this erroneous behaviour is suppressed to some degree.

Another observation is that the instrument Classic Guitar appears for the MFCCs-based method somewhat confusable with other string instruments. On the other hand, the APGD and combined methods make such confusions significantly less often, which corresponds to the previously made conclusion on the applicability of LPC-based methods for string instruments.

Additional observations can be drawn from the presented data, however, it is not always as straightforward to find a meaningful explanation to the observed phenomena. For instance, it remains unclear why MFCCs sometimes confuse Tuba with Electric Bass. However, this is overcome by the APGD-based methods, which show an almost 100% accuracy on Tuba, presumably, due to the importance of the phase characteristics of the sound of this instrument.

Table 3. Confusion matrices obtained with the evaluation of MFCCs, APGD and combined method with selecting 120 features in the 9 instrument case.

MFCCs

Instrument \ Recognised as	**Pn**	**ClG**	**ElG**	**ElB**	**Trmb**	**Tub**	**Bsn**	**Clrn**	**Bnj**
Piano	**94**	-	-	-	-	-	-	-	6
Classic Guitar	6	**77**	-	9	-	-	-	-	8
Electric Guitar	9	-	**91**	-	-	-	-	-	-
Electic Bass	-	24	-	**76**	-	-	-	-	-
Trombone	-	-	-	-	**98**	-	-	1	1
Tuba	-	-	-	7	-	**92**	1	-	-
Bassoon	3	2	1	2	4	-	**76**	2	10
Clarinet	-	-	-	2	1	-	-	**97**	1
Banjo	2	42	-	3	-	-	1	-	**52**

APGD

Instrument \ Recognised as	**Pn**	**ClG**	**ElG**	**ElB**	**Trmb**	**Tub**	**Bsn**	**Clrn**	**Bnj**
Piano	**63**	2	-	-	-	-	-	2	34
Classic Guitar	3	**95**	-	-	-	-	-	-	2
Electric Guitar	-	-	**96**	-	-	-	-	1	3
Electic Bass	1	10	-	**89**	-	-	-	-	-
Trombone	-	-	-	-	**88**	-	1	2	8
Tuba	-	-	-	1	-	**99**	-	-	-
Bassoon	-	-	-	-	4	-	**79**	13	3
Clarinet	-	3	-	1	-	-	-	**92**	5
Banjo	1	21	-	11	1	-	-	-	**66**

APGD + MFCCs + FeatSel 120

Instrument \ Recognised as	**Pn**	**ClG**	**ElG**	**ElB**	**Trmb**	**Tub**	**Bsn**	**Clrn**	**Bnj**
Piano	**81**	2	-	-	-	-	-	-	17
Classic Guitar	4	**93**	-	2	-	-	-	-	1
Electric Guitar	-	-	**99**	-	-	-	-	-	-
Electic Bass	1	21	-	**78**	-	-	-	-	-
Trombone	-	-	-	-	**92**	-	4	1	4
Tuba	-	-	-	3	-	**97**	-	-	-
Bassoon	-	-	-	1	2	-	**74**	9	14
Clarinet	-	2	-	2	-	-	-	**97**	-
Banjo	1	30	-	8	-	-	-	-	**61**

5 Conclusions

This paper studies the use of all-pole group delay features for musical instrument recognition. The proposed method of utilising the APGD feature for the given problem has shown to be valid, with its performance on the comparable levels with the commonly used MFCCs. The absolute recognition accuracy gain has shown to be up to 7% in the simpler classification scenario. In the complex classification scenario, APGD on its own shows somewhat lower performance, however, by incorporating the combined features with feature selection, accuracy gains are present in all of the evaluated cases. The work has shown that by combining the relevance of linear predictive analysis for instrument recognition with the significance of the phase information, often neglected by the commonly used features, APGD demonstrates its effectiveness and a promising potential for musical instrument recognition.

As a future research suggestion, it is worthwhile to study the performance of the proposed method in a group-wise classification scenario, as opposed to the currently presented instrument-wise case. Namely, training models of groups, composed of instruments, similar in terms of the physics of their sound production, could reveal interesting dependencies and enable a more thorough investigation of the importance of the phase for musical instrument recognition.

References

1. Agostini, G., Longari, M., Pollastri, E.: Musical instrument timbres classification with spectral features. In: Multimedia Signal Processing, 2001 IEEE Fourth Workshop on. pp. 97–102 (2001)
2. Alsteris, L.D., Paliwal, K.K.: Short-time phase spectrum in speech processing: A review and some experimental results. Digital Signal Processing 17(3), 578–616 (2007)
3. Banno, H., Lu, J., Nakamura, S., Shikano, K., Kawahara, H.: Efficient representation of short-time phase based on group delay. In: Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on. vol. 2, pp. 861–864 (May 1998)
4. Bozkurt, B., Couvreur, L., Dutoit, T.: Chirp group delay analysis of speech signals. Speech Commun. 49, 159–176 (2007)
5. Diment, A., Heittola, T., Virtanen, T.: Semi-supervised learning for musical instrument recognition. In: 21st European Signal Processing Conference 2013 (EUSIPCO 2013). Marrakech, Morocco (Sep 2013)
6. Diment, A., Padmanabhan, R., Heittola, T., Virtanen, T.: Modified group delay feature for musical instrument recognition. In: 10th International Symposium on Computer Music Multidisciplinary Research (CMMR). Marseille, France (Oct 2013)
7. Duxbury, C., Davies, M., Sandler, M.: Separation of transient information in musical audio using multiresolution analysis techniques. In: Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01). Limerick, Ireland (2001)
8. Eronen, A.: Comparison of features for musical instrument recognition. In: Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the. pp. 19–22 (2001)

9. Fletcher, N.H., Rossing, T.D.: The Physics of Musical Instruments. Springer (1998)
10. Fuhrmann, F.: Automatic musical instrument recognition from polyphonic music audio signals. Ph.D. thesis, Universitat Pompeu Fabra (2012)
11. Giannoulis, D., Klapuri, A.: Musical instrument recognition in polyphonic audio using missing feature approach. Audio, Speech, and Language Processing, IEEE Transactions on 21(9), 1805–1817 (Sept 2013)
12. Goto, M., Hashiguchi, H., Nishimura, T., Oka, R.: RWC music database: Music genre database and musical instrument sound database. In: Proc. of the 4th International Conference on Music Information Retrieval (ISMIR). pp. 229–230 (2003)
13. Hacihabiboglu, H., Canagarajah, N.: Musical instrument recognition with wavelet envelopes. In: Proc. Forum Acusticum Sevilla (CD-ROM) (2002)
14. He, X., Cai, D., Niyogi, P.: Laplacian score for feature selection. In: NIPS. vol. 186, p. 189 (2005)
15. Hegde, R., Murthy, H., Gadde, V.: Significance of the modified group delay feature in speech recognition. Audio, Speech, and Language Processing, IEEE Transactions on 15(1), 190–202 (2007)
16. Jensen, K.: Timbre Models of Musical Sounds: From the Model of One Sound to the Model of One Instrument. Report, Københavns Universitet (1999)
17. Kaminsky, I., Materka, A.: Automatic source identification of monophonic musical instrument sounds. In: Neural Networks, 1995. Proceedings., IEEE International Conference on. vol. 1, pp. 189–194. IEEE (1995)
18. Karjalainen, M., Hrm, A., Laine, U.K., Huopaniemi, J.: Warped filters and their audio applications. In: Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on. pp. 4–pp. IEEE (1997)
19. Klapuri, A.: Analysis of musical instrument sounds by source-filter-decay model. In: Acoustics, Speech and Signal Processing, 2007. IEEE Int. Conf. on. vol. 1, pp. I–53–I–56 (2007)
20. Kostek, B., Czyzewski, A.: Representing musical instrument sounds for their automatic classification. Journal of the Audio Engineering Society 49(9), 768–785 (2001)
21. Makhoul, J.: Linear prediction: A tutorial review. Proceedings of the IEEE 63(4), 561–580 (1975)
22. Marques, J., Moreno, P.J.: A study of musical instrument classification using gaussian mixture models and support vector machines. Cambridge Research Laboratory Technical Report Series CRL 4 (1999)
23. Meillier, J.L., Chaigne, A.: AR modeling of musical transients. In: Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on. pp. 3649–3652. IEEE (1991)
24. Murthy, H., Gadde, V.: The modified group delay function and its application to phoneme recognition. In: Acoustics, Speech, and Signal Processing, 2003. Proceedings (ICASSP '03). 2003 IEEE International Conference on. vol. 1, pp. I–68–71 vol.1 (2003)
25. Rajan, P., Kinnunen, T., Hanili, C., Pohjalainen, J., Alku, P.: Using group delay functions from all-pole models for speaker recognition. In: Proc. Interspeech 2013. pp. 2489–2493 (2013)
26. Sturm, B., Morvidone, M., Daudet, L.: Musical instrument identification using multiscale mel-frequency cepstral coefficients. Proc. of the European Signal Processing Conference (EUSIPCO) pp. 477–481 (2010)
27. Yegnanarayana, B.: Formant extraction from linear-prediction phase spectra. The Journal of the Acoustical Society of America 63(5), 1638–1640 (1978)