

# Towards Music Performer Recognition Using Timbre Features

Magdalena Chudy

Centre for Digital Music, School of Electronic  
Engineering and Computer Science  
Queen Mary University of London  
Mile End Road, E1 4NS London, UK  
Tel: +44 (0)20 7882 7480

magdalena.chudy@elec.qmul.ac.uk

Simon Dixon

Centre for Digital Music, School of Electronic  
Engineering and Computer Science  
Queen Mary University of London  
Mile End Road, E1 4NS London, UK  
Tel: +44 (0)20 7882 7681

simon.dixon@elec.qmul.ac.uk

## ABSTRACT

In this study, we investigate whether timbre descriptors commonly used for instrument recognition can serve as discriminators between different players performing on the same instrument. To address the problem we compare timbre features extracted from monophonic recordings of six cellists playing an excerpt from Bach's *1<sup>st</sup> Cello Suite* on two different cellos. We test each descriptor's ability to reflect timbre differences between players and evaluate its adequacy for classification using standard analysis of variance.

## Keywords

Timbre dissimilarities, performer discrimination, timbre descriptors

## 1. INTRODUCTION

A classical music performer interprets a musical piece using parameters such as dynamics, tempo, articulation, timing and timbre, which are essential to enliven and shape an objective musical score into an emotionally expressive performance. They form an individual playing style that is heard in any musical interpretation. To trace and capture performer stylistic features by measuring performance aspects in musical recordings is still an open problem [17, 20].

### 1.1 Recognising performers from their playing style

Several previous works have demonstrated the possibility of distinguishing music performers by measuring variations in tempo, dynamics, articulation and timing [3, 12, 18, 19, 22]. Different methods for feature extraction were proposed as well various approaches to the task of modelling "expressivity" patterns, starting with statistical analysis and machine learning techniques.

For example Widmer et al. [16, 21, 23] explored statistical data

derived from dynamics (loudness) and tempo changes, extracted using the BeatRoot [4] system, to retrieve expressive features. For performance measuring and analysis, twelve fragments from Mozart's *Piano Sonatas* recorded by six famous artists were chosen. In the first two studies they tested six different machine learning algorithms for a classification task in a pair-wise discrimination setting. The obtained overall identification accuracies varied between 60 and 70%.

Using segmentation and clustering techniques, Saunders et al. [16] analysed the same Mozart sonata recordings mentioned above, and obtained a set of prototypes from "performance worms" [5]. Each prototype was then associated with a letter to create a "general performance alphabet". Subsequently, performance trajectories were represented by strings of letters. To identify performers in pairs, they exploited String Kernels and Support Vector Machine methods with accuracy between 75 and 81%.

Ramirez et al. [14] used symbolic descriptions extracted from monophonic recordings to recognise three jazz saxophonists from their playing style. Several jazz standards were performed by each player and then represented by sets of note-level descriptors. Every individual note was described by its intra-note features corresponding to the perceptual (timbre) characteristics (based on instantaneous energy and fundamental frequency) and inter-note features representing the melodic context. Machine learning algorithms were applied to cluster similar (in terms of timbre) individual notes into performance alphabets and link them with the players. The average classification accuracy reported was 97.03% for short phrase comparisons and 96.77% for long phrases. In his next experiments Ramirez et al. [15] employed a very similar approach for identification of music performers from violin recordings. They analysed nine Celtic jigs performed by two violinists and obtained 100% classification accuracy.

Despite promising outcomes, these studies revealed that "expressivity" patterns are difficult to generalize, since all expressive parameters, including timbre, can vary from one performance to another and may depend on an interpreted musical piece, e.g. being determined by different composers' styles or musical genres. However, among the stylistic features, timbre predominates as the feature most dependent on a performer's unique physical and perceptual abilities. On each instrument, a performer creates an individual set of timbre colours, a timbral "fingerprint". If individual timbre features are able to characterize a performer, then timbral dissimilarities can be used for performer discrimination. This concept of



distinguishing performers has one serious drawback. Timbre features extracted from digital recordings are influenced by the acoustical conditions of a recording session which affect the quality and spectral content of the music signal. Consequently, the results of timbre analysis will be biased by recording artefacts.

Music performer timbre-based recognition was proposed in [2]. The study focused on a performer's timbre created on an instrument which is unique to every musician. Linear Prediction Cepstral Coefficients (LPCCs) were exploited to capture performers' unique timbre characteristics. LPCC parameters, capable of modelling the overall spectral envelope, were extracted from experimental studio recordings made by six cello players on two different cellos. Vector quantization was employed for building performers' models and classification. The total recognition accuracy, obtained from experiments performed on each cello separately and two instruments combined together, was above 88% in all cases. This preliminary study confirmed that performer timbre characteristics cannot be easily separated from timbre characteristics of an instrument, and therefore additional investigation is necessary to explain which spectral, temporal or spectrotemporal features depend specifically on the performer's manner of playing.

## 1.2 Describing timbre

Searching for a comprehensive model of timbre describing different instrument sounds, a variety of acoustic descriptors were developed based on spectral, temporal and spectro-temporal sound characteristics. The majority of timbre descriptors were found in correlation with human perception of sound thanks to extensive perceptual studies [7, 9, 11]. Their standardised definitions are incorporated into the MPEG-7 standard for audio data representation [8]. For instrument description and classification purposes, a feature extraction stage often includes calculating temporal descriptors e.g. attack time, decay time or temporal centroid; spectral descriptors e.g. spectral centroid, spectral deviation, spectral spread, irregularity, tristimuli, odd/even ratio, amplitude envelope; and spectrotemporal descriptors e.g. spectral flux, roughness. Various combination of these features were used with success for instrument sound discrimination.

In this study, we aim to find out which descriptors can be correlated with the performer's manner of playing affecting perceived timbre of an instrument, and to test whether these descriptors can help to discern various players.

## 2. METHOD

In order to examine if there are salient differences of timbre between different performers playing the same instrument, we analysed monophonic recordings and extracted a set of timbre descriptors chosen amongst those mentioned in the previous section. The subsequent steps of audio data processing and timbre descriptor computation are presented in detail in the following sections.

### 2.1 Sound corpus

If timbre descriptors are able to show systematic dissimilarities between performers, to obtain statistically significant results, it would be ideal to use the same music excerpts played multiple

times by each performer on a few instruments belonging to the same class. As such a database was not available at the time of this study, we exploited instead a set of studio recordings made by six cello players on two different cellos, as used in the author's previous experiments [2]. An excerpt of the *Gigue* from Bach's 1<sup>st</sup> Cello Suite was recorded twice by each player on each instrument (i.e. 24 recordings in total). All recordings were taken from the near-field, at a distance of approximately 1m, but the exact settings and detailed specification of recording equipment were not provided. The music signals were captured in mono channel .wav format at a sampling rate of 44.1 kHz and 16 bits per sample.

### 2.2 Note extraction

Instead of analysing the entire music sequence we concentrated on timbre descriptors at the note-level. From the first two motives of the *Gigue* we extracted six notes of the same pitch D3 (fundamental frequency = 146.83 Hz) occurring in different positions within a musical phrase (see Figure 1). The positions were defined as anacrusis<sup>1</sup> (A), transitional (T), upbeat<sup>2</sup> (U) and downbeat<sup>3</sup> (D). This was intended to ensure timbre variations between notes even if performed with similar articulation.

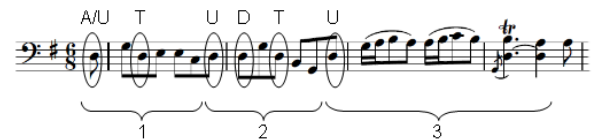


Figure 1. *Gigue* from Bach's 1<sup>st</sup> Cello Suite, bars 1-4. Curly brackets indicate the musical phrase structure. Annotations above selected notes denote their position in a melody and/or measure (A – anacrusis, T – transitional, U – upbeat, D – downbeat)

We used a note onset detector proposed by [6] with some manual corrections of the results. We repeated the procedure across recordings to obtain audio data consisting of 24 notes for each cello player (12 notes on each cello). The length of audio signals varied between 139-417 ms.

### 2.3 Sound representation

For audio data analysis we applied the sinusoidal plus residual model of sound proposed by Serra [24] as a part of the spectral modelling synthesis (SMS) framework. The audio signal is decomposed here into a sum of sinusoids with time-varying amplitudes and frequencies (corresponding to the stable partials of a sound) and a residual component (representing the noise part). The definition of the sound  $s(t)$  is given by:

$$s(t) = \sum_{k=1}^N A_k(t) \cos [q(t)] + e(t), \quad (1)$$

<sup>1</sup> anacrusis – the initial note/notes of a melody that precede the first full bar

<sup>2</sup> upbeat – the last beat of a measure, a weak beat, which anticipates the downbeat

<sup>3</sup> downbeat – the first beat of a measure

**Table 1. Timbre descriptors and their parameters**

Descriptor	Attack time			Steady state		
HSC	HSC_at_mean	HSC_at_std	HSC_at_range	HSC_st_mean	HSC_st_std	HSC_st_range
HSI	HSI_at_mean	HSI_at_std	HSI_at_range	HSI_st_mean	HSI_st_std	HSI_st_range
OEHR	OEHR_at_mean	OEHR_at_std	OEHR_at_range	OEHR_st_mean	OEHR_st_std	OEHR_st_range
TRI1	TRI1_at_mean	TRI1_at_std	TRI1_at_range	TRI1_st_mean	TRI1_st_std	TRI1_st_range
TRI2	TRI2_at_mean	TRI2_at_std	TRI2_at_range	TRI2_st_mean	TRI2_st_std	TRI2_st_range
TRI3	TRI3_at_mean	TRI3_at_std	TRI3_at_range	TRI3_st_mean	TRI3_st_std	TRI3_st_range
HRMS	HRMS_at_mean	HRMS_at_std	HRMS_at_range	HRMS_st_mean	HRMS_st_std	HRMS_st_range
AT						

where  $A_k(t)$  and  $q_k(t)$  are the instantaneous amplitude and phase of the  $k^{\text{th}}$  sinusoid,  $N$  is the number of sinusoids, and  $e(t)$  is the noise component at time  $t$  (in seconds). This representation method reflects the additive nature of musical sounds composed of the fundamental frequency and its harmonics, and enables to track short time fluctuations of the amplitude of each harmonic.

We chose to analyse 50 tracks of harmonics including the fundamental. We used STFT frame-based analysis with a Blackman-Harris window of 2048-sample size and 87.5% overlap which gave us a 5.8 ms time resolution. The length of the FFT was set to 4096 points (2048 plus zero-padding) resulting in a 10.76 Hz frequency resolution. The minimum amplitude value was set at a level of -100 dB.

## 2.4 Timbre descriptors

After the audio processing stage each note in our dataset was represented by a matrix of 50 harmonics with their instantaneous amplitudes (in dB) captured at time instants  $t$ . For the following calculations we did not consider the residual parts of sounds, which we leave for the next round of experiments. From this data, at each time point  $t$ , we computed seven timbre descriptors chosen to capture timbre differences between players: the harmonic spectral centroid, the harmonic spectral irregularity, the odd/even harmonic ratio, the first, second and third tristimulus, and the harmonic RMS energy. The next paragraph provides definitions and formulas for each descriptor. In the following equations  $A_k(t)$  denotes the amplitude of the  $k^{\text{th}}$  harmonic and  $N$  is the number of harmonics.

The Harmonic Spectral Centroid (HSC) can be defined as the amplitude weighted mean of the harmonic peaks of the spectrum, and is correlated with the perceptual brightness of a sound:

$$HSC(t) = \frac{\sum_{k=1}^N k A_k(t)^2}{\sum_{k=1}^N A_k(t)^2}, \quad (2)$$

The Harmonic Spectral Irregularity (HSI) measure proposed by Jensen [10] calculates amplitude differences between subsequent odd and even harmonics:

$$HSI(t) = \frac{\sum_{k=1}^N [A_k(t) - A_{k+1}(t)]^2}{\sum_{k=1}^N A_k(t)^2}, \quad (3)$$

The Odd/Even Harmonic Ratio (OEHR) describes the relation between the energy content of odd and even harmonics:

$$OEHR(t) = \frac{\sum_{k=2}^{N/2} A_{2k-1}(t)^2}{\sum_{k=1}^{N/2} A_{2k}(t)^2}, \quad (4)$$

Tristimulus (TRI) descriptors indicate the energy content in groups of harmonics and are defined as follows:

$$TRI1(t) = A_1(t)^2 \sqrt{\sum_{k=1}^N A_k(t)^2}, \quad (5)$$

$$TRI2(t) = \sum_{k=2}^4 A_k(t)^2 \sqrt{\sum_{k=1}^N A_k(t)^2}, \quad (6)$$

$$TRI3(t) = \sum_{k=5}^N A_k(t)^2 \sqrt{\sum_{k=1}^N A_k(t)^2}, \quad (7)$$

The Harmonic RMS (HRMS) corresponds to the amplitude envelope of harmonics and is often used to show fluctuations in spectral energy over time:

$$HRMS(t) = \sqrt{\frac{\sum_{k=1}^N A_k(t)^2}{N}}, \quad (8)$$

In addition to spectral descriptors we calculated the Attack Time (AT) of each note. Multiple perceptual studies of timbre have found this parameter to be one of the most important dimensions in timbre space, and it has been shown to be applicable for instrument classification [10, 11]. We used the RMS energy of each signal frame to obtain the amplitude envelope and set amplitude thresholds at 10% and 90% of the maximum value to find the time points of the start and the end of each attack. A similar approach was applied in [1].

## 2.5 Time averaging of timbre descriptors

As the recorded cellists were allowed to interpret music without any constraints regarding tempo and timing, the durations of extracted notes varied across players. There were also duration differences occurring in connection with the note positions in a phrase. To be able to compare the descriptors at the note-level we divided each note into two parts according to the attack/rest envelope model proposed by Peeters [13]. We used time instants previously obtained for the Attack Time parameter as limiters for the attack part of a note and conceded the remaining frames as the rest, namely the steady state part. For every note in the dataset we calculated the mean of each descriptor, having it represented by two values derived from the attack and steady state part respectively. In addition, to characterise the descriptors' variations within the duration of the attack and steady state we computed their standard deviation and range values. Table 1 presents the labels of all parameters computed for each single note in the dataset.

## 3. RESULTS

### 3.1. Player factor

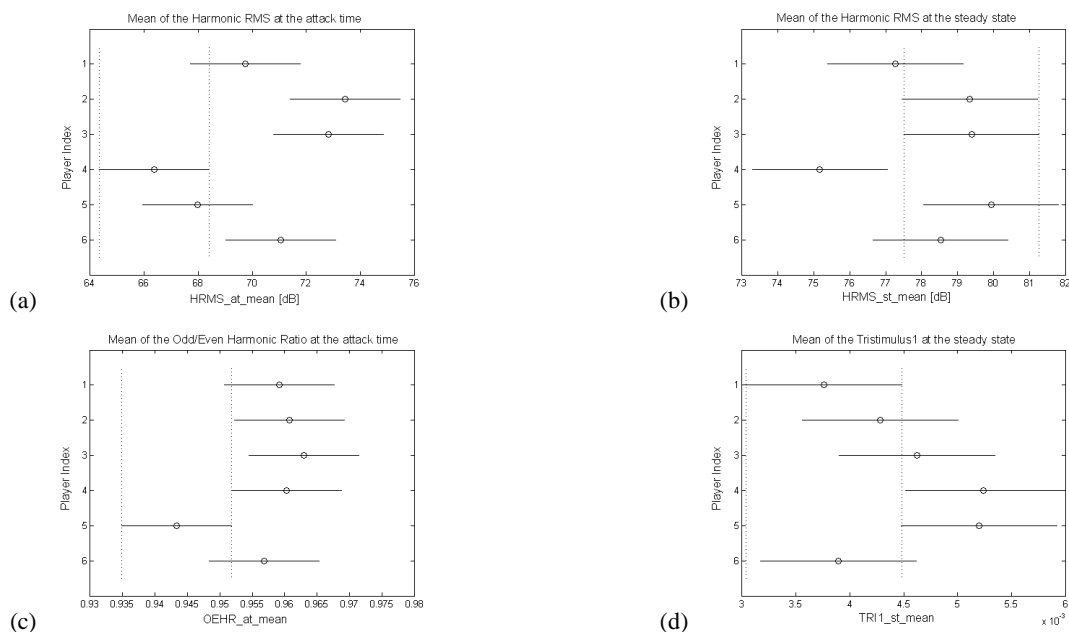
In order to test for each descriptor whether any differences exist between players, and whether these differences are consistent when an instrument changes, we employed a two-way analysis of variance (two-way ANOVA). We set the player and the cello as

**Table 2. Parameters with a significant player-effect**

Descriptor	OEHR_at_mean	HRMS_at_mean	HRMS_st_mean	HRMS_st_std	HRMS_st_range	TRI1_st_mean
p-value	0.0184	0.0000	0.0041	0.0212	0.0195	0.0102

**Table 3. Parameters with a significant instrument-effect**

Descriptor	HSC_at_mean	HSC_st_mean	OEHR_st_std	OEHR_st_range	TRI1_st_std	TRI2_at_mean	TRI2_st_mean	TRI3_st_mean
p-value	0.0000	0.0000	0.0068	0.0025	0.0411	0.0287	0.0054	0.0333



**Figure 2. The comparison of means between players (circles) for the four parameters with the lowest p-values. Horizontal lines denote 95% confidence intervals**

independent factors and examined their influence on each parameter defined in Table 1. If a player-effect on a descriptor was observed, i.e. a critical p-value was below 0.05, it meant that a significant difference in means over all 24 notes occurred at least in one pair of players. In that case we applied a multiple comparison test to determine which pairs of means were significantly different. Descriptors affected by a player factor are listed in Table 2 and the results of comparison tests for the four parameters with the lowest p-values are shown in Figure 2.

To confirm whether our results are similar considering only one instrument, we performed one-way ANOVA experiments using notes from the first cello dataset. This analysis yielded the identical set of parameters dependent on a player factor as presented in Table 2.

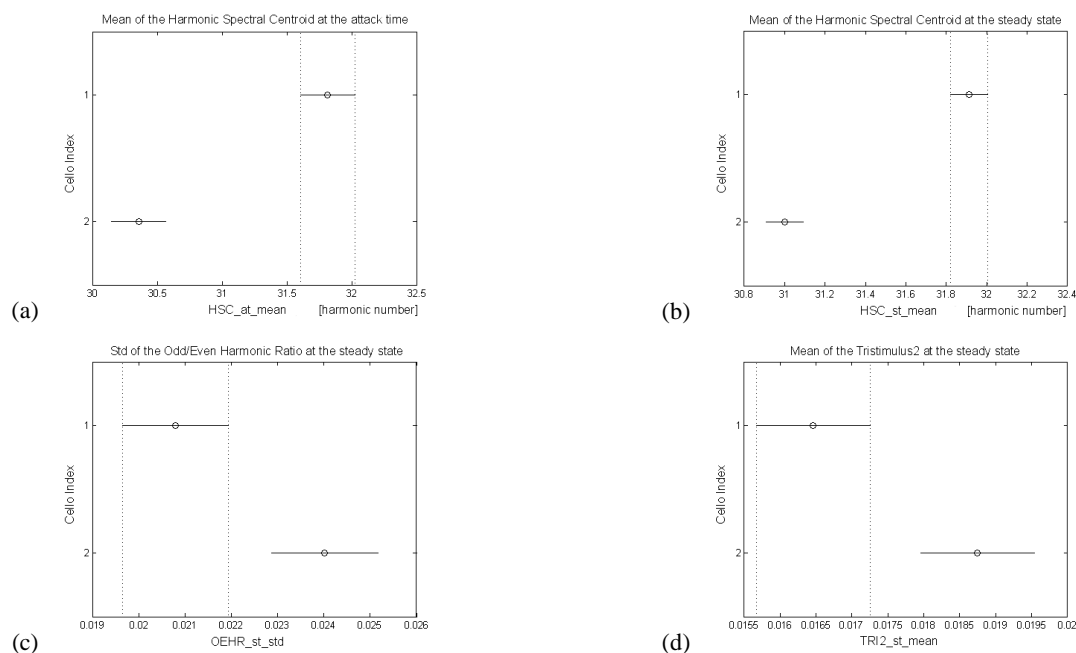
Figure 2a and 2b shows the mean values of the harmonic RMS energy computed for each player across 24 notes at their attack and steady parts respectively. With the assumption that dynamic levels during the recording session were exactly the same, this parameter should reflect the intensity patterns of the players. Indeed, for all players the values of HRMS are lower during the attack time (from approx. 66 up to 74 dB) and much higher during the steady state (from approx. 75 up to 80 dB). This is due to the fact that the steady part contains the maximum energy value of the note. One can notice that Player 4 has in general the lowest energy values in both cases. The individual differences between the two time points vary from 5.5 dB for Player 2 up to

12 dB for Player 5. This may imply that in terms of musical expressiveness notes performed by Player 2 are less dynamically shaped and developed over time in comparison to Player 5.

When testing means for differences between pairs of players, one should consider the size of confidence intervals for each mean value. As shown in Figure 2, two means are significantly different if their intervals are disjoint, and are not significantly different if their intervals overlap. Therefore, based on the spectral energy at the attack time, we can distinguish Player 4 from Player 2, 3 and 6, or Player 5 from Player 2 and 3, but Player 1 is indistinguishable from the others.

Analysing the Odd/Even Ratio descriptor (Figure 2c) one can see that Player 5 has slightly lower content of odd harmonics in the attack portions of tones. The rest of the players have this parameter values practically at the same level and cannot be distinguish from each other. It is worth noticing that for this parameter we observed the strong interaction of the two independent factors (p-value = 0.007). This can be explained by the fact that the odd/even harmonic proportions in the spectrum are highly determined by the individual characteristics of each cello and the player can control them only to some extent.

The mean values of the first Tristimulus computed for each player across 24 notes at their steady parts are presented in Figure 2d. This descriptor corresponds to the energy content of the first harmonic, i.e. the fundamental, measured in proportion to the overall spectral energy. It is evident that even though the means



**Figure 3. The comparison of means between cellos (circles) for the four parameters with the lowest p-values. Horizontal lines denote 95% confidence intervals**

vary between players the wide confidence intervals overlap. This implies that we can not easily discern players based on this parameter.

### 3.2. Instrument factor

In addition to the parameters with a significant player-effect, the two-way ANOVA analysis also indicated descriptors dependent on the instrument factor. Table 3 presents the parameters and the obtained p-values. The results of comparison tests for the four parameters with the lowest p-values are shown in Figure 3.

The mean values of the Harmonic Spectral Centroid calculated for each cello at the attack and steady parts of notes (Figure 3a and 3b respectively) are well separated from each other. Cello 1 has the spectral centroid located close to the 32<sup>nd</sup> harmonic (associated frequency in Hz depends on the fundamental frequency), while for Cello 2, the spectral centroid varies between the 30<sup>th</sup> harmonic at the attack time and the 31<sup>st</sup> harmonic at the steady state. These results confirm the correlation existing between the spectral centroid and timbre characteristics of instruments. The player performing on each cello had no effect on the descriptor values. Since the spectral centroid depends on the fundamental frequency, in order to differentiate between the two instruments the descriptor must be normalised by the fundamental frequency of each compared tone to avoid biased results.

Comparing the results from Table 2 and Table 3 for the Tristimulus descriptors, we observed that the first Tristimulus derived from the steady portions of tones showed to be influenced by a player, with some variations depending on a cello (TRI1\_st\_std p-value = 0.0411), while the second Tristimulus was found to be affected by an instrument over the whole durations of tones. It can be explained by the fact that a player has more impact on the energy content of the fundamental, while the content of 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> harmonics in the spectrum is strongly determined by an instrument. Also the third Tristimulus that

corresponds to the content of higher harmonics in the spectrum (from 5<sup>th</sup> to 50<sup>th</sup> in our case), measured during the steady parts of notes, showed to be dependent, to some extent, on an instrument (p-value = 0.0333). Looking at Figure 3d, one can see that the mean values of the second Tristimulus for each cello are well separated, thus easy to discriminate.

Surprisingly, the mean values of Odd/Even Harmonic Ratio obtained from the steady parts of tones showed no dependency on instrument. Instead, its standard deviation occurred affected by the factor (p-value = 0.0068 for std and p-value = 0.0025 for range) indicating that Cello 1 has more stable proportions of odd and even harmonics during the steady portions of notes then Cello 2 (see Figure 3c), and this can be used to differentiate between the two instruments.

## 4. DISCUSSION

We have to admit that these preliminary results are somewhat surprising. In fact, amongst all spectral descriptors we chose for the experiments, the harmonic amplitude envelope (HRMS) proved to be the most sensitive to the player factor. The rest of descriptors demonstrated some player-effect (the Odd/Even Ratio and the first Tristimulus, but not very significant) or no player influence at all (the Irregularity, the Spectral Centroid, and the second and third Tristimulus). The first explanation that comes up is that we intentionally explored only sinusoidal parts of audio signals and did not consider the noise components and this might cause the lack of important spectral information about individual timbres. This led us to the conclusion that in the next experiments we need to subsume the residual parts of signals into descriptor calculation process.

Detailed collation of the HRMS values at the attack and steady parts of notes in pairs of players showed that differences may not appear necessarily at the same time points but between them, hence indicating how the parameter evolves over time. Such a

significant dissimilarity was observed between Player 2 whose mean dynamic range of tones amounted to 5.5 dB and Player 5 whose mean dynamic range of tones amounted to 12 dB. This type of comparison can be applied to the rest of parameters and may yield interesting results.

Worth noticing was that the Attack Time, the parameter reported as an important coordinate of timbre spaces, showed to be useless in our experiments. This may be due to the simple method of computation (the alternative adaptive threshold method can be used [13]) or the scale we used to capture it (linear instead of logarithmic), or to the fact that this parameter is not adequate for our purpose.

The ANOVA analysis of the cello factor produced more promising outcomes. It proved the ability of the descriptors to discern between the two cellos with the exceptional emphasis on the Harmonic Spectral Centroid performance. This is not a surprise as they were developed to characterise various instrument sounds and their initial application was to distinguish instrument timbres.

## 5. CONCLUSIONS

In our first round of experiments, we analysed and compared a set of timbre descriptors using the two-way ANOVA procedure. We focused on testing each timbre descriptor separately for its ability to discern cello players. The obtained results showed that three of the descriptors are correlated with the player factor and amongst them only HRMS demonstrated stronger dependency. They also pointed out parameters related to the instrument factor, already reported as good discriminators between instruments.

It must be highlighted here that differences in timbre observed between players on separate parameters were not sufficient to enable player discrimination based on a single descriptor. This was obviously due to the complex and multidimensional nature of musical timbre which could not be explained by the simple linear model of data analysis employed here. If our aim is to effectively capture timbre dissimilarities between players we need to apply a large set of timbre descriptors in combination with non-linear multivariate methods of data modelling, such as multidimensional scaling and principal component analysis, in our next experiments.

## REFERENCES

- [1] Barthelet, M. *et al.* Acoustical correlates of timbre and expressiveness in clarinet performance. *Music Perception*, 2010, accepted for publication.
- [2] Chudy, M. Automatic identification of music performer using the linear prediction cepstral coefficients method. *Archives of Acoustics*, 33, 1, 2008, 27-33.
- [3] Dillon, R. *On the recognition of expressive intentions in music playing: a computational approach with experiments and applications*. PhD Thesis, InfoMus Lab, DIST, University of Genoa, 2004.
- [4] Dixon, S. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30, 1, 2001, 39-58.
- [5] Dixon, S. *et al.* The Performance Worm: Real time visualisation of expression based on Langner's tempo-loudness animation. In *Proceedings of the International Computer Music Conference (ICMC '02)*, 2002, 361-364.
- [6] Duxbury, C. *et al.* Complex domain onset detection for musical signals. In *Proceedings of the 6th Conference on Digital Audio Effects (DAFx '03)*, 2003.
- [7] Grey, J. M. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61, 1977, 1270-1277.
- [8] ISO/IEC 15938-4. *Information technology – Multimedia content description interface – Audio*. 2002.
- [9] Iverson, P. and Krumhansl, C. L. Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, 94, 1993, 2595-2603.
- [10] Jensen, K. *Timbre models of musical sounds*. PhD Thesis, Department of Computer Science, University of Copenhagen, 1999.
- [11] McAdams, S. *et al.* Perceptual scaling of synthesized musical timbres – common dimensions, specificities, and latent subject classes. *Psychological Research*, 58, 1995, 177-192.
- [12] Molina-Solana, M. *et al.* Using expressive trends for identifying violin performers. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR '08)*, 2008, 495-500.
- [13] Peeters, G. *A large set of audio features for sound description (similarity and description) in the CUIDADO project*. I.R.C.A.M. Technical Report version 1.0, Paris, 2004.
- [14] Ramirez, R. *et al.* Identifying saxophonists from their playing styles. In *Proceedings of the 30th AES International Conference*, 2007.
- [15] Ramirez, R. *et al.* Performer identification in celtic violin recordings. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR '08)*, 2008, 483-488.
- [16] Saunders, C. *et al.* Using string kernels to identify famous performers from their playing style. In *Proceedings of the 15th European Conference on Machine Learning*, 2004, 384-395.
- [17] Sloboda, J. A. Individual differences in music performance. *Trends in Cognitive Sciences*, 4, 2000, 397-403.
- [18] Stamatos, E. and Widmer, G. Automatic identification of music performers with learning ensembles. *Artificial Intelligence*, 165, 2005, 37-56.
- [19] Tobudic, A. and Widmer, G. Learning to play like the great pianists. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI '05)*, 2005.
- [20] Widmer, G. and Goebel, W. Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, 33, 2004, 203-216.
- [21] Widmer, G. and Zanon, P. Automatic recognition of famous artists by machine. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI '04)*, 110, 2004.
- [22] Widmer, G. *et al.* In search of the Horowitz factor. *AI Magazine*, 24, 2003, 111-130.
- [23] Zanon, P. and Widmer, G. Recognition of famous pianists using machine learning algorithms: first experimental results. In *Proceedings of the 14th Colloquium on Musical Informatics (CIM '03)*, 2003.
- [24] Zölzer, U. *DAFX – Digital Audio Effects*. Wiley, Chichester, 2002, Chapter 10.