Features detecting bad violin playing

Jane Charles*, Derry Fitzgerald*, Eugene Coyle*
School of Electrical Engineering Systems

Dublin Institute of Technology

email : janecharles@gmx.com

derry.fitzgerald@dit.ie

eugene.coyle@dit.ie

Abstract— The research presented in this paper looks at the violin's entire timbre space and how it is influenced by the player. It differs from existing violin sound analyses research in that it includes beginner player sound. Three features which are accurate representational measures from which a computer can determine a violinist's playing standard are detailed in this paper. These features analyse content which falls into the frequency region below the normal playing range of the violin. This means below G3 which is approximately 196 Hz when a violin is tuned to A440. Research influencing this work comes from many areas including violin acoustics, music teaching methods and aids, music information retrieval, automatic accompaniment systems, speech recognition and player-instrument relationships. Given that much existing work has focused on how to best represent violin sound for use within instrument recognition systems [1, 2] and also from an instrument quality perspective [3], little reflects or represents the relationship between player and sound produced. The outcomes of this research can be applied in various systems including the development of a violin or bowed string instrument teaching aid.

***Keywords* – violin timbre, CQT, PSD, SCM**.

I INTRODUCTION

The objective of this work is to find suitable features which can be used to capture and represent violin timbre change as influenced by a player. The ultimate aim being to use these features in a classification system such as could form part of a computer based violin teaching program. In the following sections, existing research is briefly presented, followed by a description of the dataset, including why and how it was obtained. The features reflecting a violinist's playing standard are detailed in Section 2 and are based on the constant Q transform (CQT) specific frequency bin information, mean power spectral density (PSD) below the violin's playing range and spectral contrast measure (SCM) filter specific content.

II EXISTING RESEARCH

Much existing research on violins has been carried out to better understand and to emulate the making of top sounding instruments, such as those made by the old, Italian master luthiers, such as Stradivarius. Many methods have been applied to gain insight into the complex interactions between the various components of stringed instruments [3] and work is ongoing considering the problem of quantifying perception relating to violin sound quality [4]. Much of this work though does not focus on playing technique and its effects on sound produced. Although the violin is the most uniform of the stringed instruments [5], much remains unknown about its acoustics. The complexities of how the violin resonates make it extremely difficult to develop a complete physical model. Work towards developing a physical model of a bowed violin string has been done [6] and Wilson has tried extracting violin performance information necessary to drive a digital waveguide model of a bowed violin [7]. With the violin, minute changes such as moving the sound post less than a millimetre greatly influence the instrument's sound [8]. Such variables, of which there are many, need to be captured by a physical model. However, information relating to physical models for various wind instruments such as the trumpet, trombone, saxophone [9], oboe [10] and piano [11] has been published.

Spectral features have been used for musical instrument timbre classification [12] as have cepstral and temporal features [1]. In these works, instruments including the violin are represented by multiple features. Features used for identifying individual instruments focus on good instrument sound and not on representing change within an instrument's timbre space. The classification of three common violin bow strokes has been done using data collected from an electric violin and a carbon bow to which sensors have been attached [13]. These works consider measurements obtained via sensors for good playing sound or advanced playing technique only. Work exploring the effect a player has on the violin sound produced is limited and there seems to be no work conducted on analysing poor violin playing technique. Finding features which are suitable for quantifying the violin's timbre space involves exploring the effect of a player on sound quality. Many features, although very useful in determining one instrument from another [1, 2], are not appropriate for representing the subtleties due to playing technique or for use within an individual instrument's timbre space.

III DATASET

Many commercial instrument sound sample libraries are available (Vienna Symphonic Libraries, London Symphony Orchestra). The Real World Computing Music Database also provides instrumental samples including violin samples as do the Electronic Music Studios at the University of Iowa. As with seemingly all sample libraries readily available for download, individual violin legato notes are included but no beginner note samples are available. As no suitable dataset was readily available encompassing both professional standard playing legato and beginner note samples, one had to be made. Much thought was given in creating this dataset in terms of what was needed, obtainable and viable. The first bow stroke a beginner must learn is *legato*, which literally means 'tied together' [14] as opposed to slurred, which refers to multiple notes in a same bow stroke. Although legato playing encompasses all lengths of bow stroke, in this work it means using full bows. Mastering this ensures enough bow control upon which the student can develop other bow strokes. Since the style or type of bow stroke used effects the readings obtained, only professional standard player *legato* notes will be used and the beginner notes will be compared to these.

The dataset consists of two same sized groups, one with beginner note samples and the other with professional standard player legato note samples, totaling 176 samples. The samples all contain one note which are of varying lengths and pitches, making it more appropriate to use features which do not dependent on ether note length or pitch. The pitch range of the dataset is any note which is played in the first position, i.e. G3 to B5. The dataset's samples were obtained using the best microphone available in a recording studio with a very dead acoustic. A Beyerdynamic M201TG dynamic microphone with hypercardioid polar pattern was used and placed as close as possible to the f-holes without disturbing the bow arm. The track was recorded onto DAT and saved as monophonic 16-bit, 44.1 kHz format wav samples. The recordings were all made in the same studio, using the same microphone and set up as well as the same violin and bow. Two professional standard players and three beginner players recorded samples.

An old French violin was used with a modern, 60g well-balanced bow. It is a relatively large violin which speaks easily and evenly throughout its frequency range and has a big, clear sound. It is an instrument that a beginner is able to play easily. At the time of these recordings, the strings on the instrument were Thomastik Dominant Mittel for the G, D and A strings and a Pirastro Oliv E string.

Listening tests have been carried out on the dataset to remove the subjectivity and to have labels associated with each sample. 21 professional string players completed the tests which in part involved labeling each sample as a beginner or as a professional player so that these labels could be contrasted with results obtained.

IV FEATURES

Certain features from the temporal, spectral and cepstral domains have been shown to differentiate successfully between the different player groups in the dataset [15]. Included in these are the waveform amplitude mean, moving mean variance and kurtosis values taken from the temporal domain. Spectral analysis permits the component frequencies present in a sound sample to be observed, giving insight into its harmonic structure and timbre. From the spectral domain, taking the spectral flatness measure (SFM) mean and the SFM variance returned effective results for distinguishing between beginner and professional standard player samples in the dataset. The real cepstral first coefficient (RCC0), RCC1, RCCs mean, RCCs variance and the RCCs kurtosis measures also served well at reflecting player standard [15]. Below, three spectrum based features are presented.

a) CQT Specific Frequency Bins

The CQT, as introduced by Brown [16], yields a log-frequency scaled time-frequency representation of the signal and is effective for visualising and exploiting information about the harmonic content of a note due to the frequency resolution. Eighth tone spacing has been selected in this work over the more often used quarter tone spacing [16] to access more information in the beginner note samples. On the violin, a quarter tone is a playable interval. The first frequency bin centre has been set at 110 Hz, well below the lowest note on the violin G3, which is approximately 196 Hz when tuned to A440. An upper frequency limit of 10 kHz has been applied. The beginner players' notes are not as cleanly executed as the professional standard legato ones. This is reflected by the presence of additional frequencies unrelated to those of the actual note as well as the harmonics not being as well defined as those displayed by the professional standard legato note samples. From observing the CQT representations of the dataset samples, differences between the beginner and professional standard legato note samples are visible.

Frequency Bin No.	*f_c (Hz)*
4	114.87
5	116.54
6	118.24
7	119.96
8	121.70
9	123.47
10	125.27
11	127.09
20	144.73

Table 1: CQT Frequency Bin Centre Frequencies which Effectively Group Beginner and Professional Standard Legato Note Sample.

The frequency content present below the playing range of the violin was investigated in the belief that unwanted frequency content reflecting playing quality may be present. Focusing within this frequency range and recalling that eighth tone spacing has been used in this study means that bin 41 has a centre frequency of 196 Hz and bin 40, that of 193.2 Hz. The mean content of each sample for the first 40 individual frequency bins was taken and revealed useful information. Separation between the dataset's professional standard and beginner player samples can be achieved using this information. To glean further insight, the content of each of these frequency bins have been looked at. Nine out of these 40 frequency bins, labelled according to their centre frequencies, are detailed in Table 1, correctly group these two player types in the dataset.

To illustrate the level of separation between the different player groups, only the mean frequency content present from frequency bins six, seven and 20 are plotted in Figure 1, where the beginner note samples are represented by a dotted line and the professional standard legato ones, by a solid one. Applying a t-test with a 0.01 significance level to the results displayed in Figure 1 returned results that are statistically significant. The null hypothesis is rejected in all three cases with p-values of $3.5*10^{-101}$, $2.1*10^{-106}$ and $3.4*10^{-80}$ respectively.

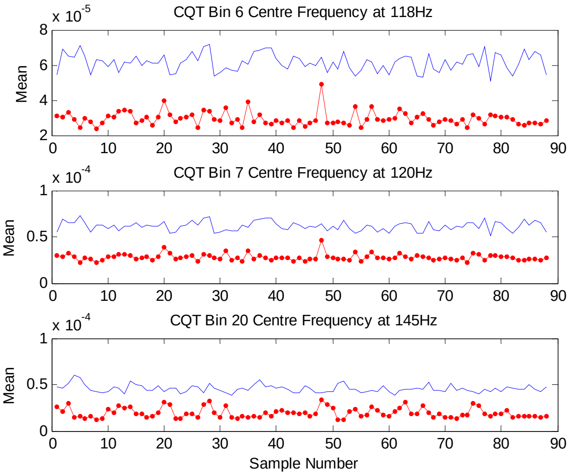

Figure 1: CQT frequency bins numbered 6 (f_c=118Hz), 7 (f_c=120Hz) and 20 (f_c=145Hz).

The professional standard legato note samples have higher mean frequency content in these frequency bins than the beginner ones do. An explanation for this gap in frequency content between the beginner and professional standard legato note samples is the excitation of instrument resonances and modes by the players. A violin's fundamental cavity resonance is at approximately 260-290 Hz [3], the frequency range for a sub-harmonic is approximately 135-145 Hz. Taking the mean frequency con-tent of the twentieth frequency bin, which has a centre frequency of 144.7 Hz, completely separates the beginner from the professional standard legato note samples in the dataset. A plausible explanation for these results is how the violin's fundamental cavity resonance at approximately 290 Hz is excited by the different players. The professional standard legato note samples all have much higher frequency content in this bin than the beginner note samples do as shown in the lower image in Figure 1. Consulting Marshall's work on violin modes in [17], the frequency content present in frequency bins six and seven, reflects modes 5 and 1 respectively. Mode 5, the first vertical cantilever of the finger-board, is at 236.5 Hz which is approximately twice the centre frequency of bin six. Mode 1 at 119.5 Hz, is the vertical reflection of the tailpiece, is reflected by frequency bin seven. The frequency content in the remaining bins may reflect specific violin modes too but this was not revealed in the material referred to. A professional standard player is expected to excite these modes more consistently and to a greater extent than a beginner player would. This accounts for the gap in average frequency content in these bins. The frequency content present in nine CQT frequency bins with centre frequencies below the lowest note on the violin provide effective discriminators between the dataset's beginner and professional standard player legato note samples.

b) Mean PSD Below 190Hz

The results obtained from certain frequency bins within the 110-190 Hz range in the CQT representations prompted researching the energy below the violin's playing frequency range. This PSD based feature presented uses the frequency content present below the violin's lowest note. Beginner notes tend to be less clear sounding due to bowing difficulties, i.e. scraping and crunching. Taking the power associated with the frequencies that fall below the violin's frequency range reflects this information. The PSD below 190 Hz feature takes the mean of the power distribution with respect to frequency present up to 190 Hz. The mean power present below 190 Hz in each sample is displayed in Figure 2. The statistical significance of these results has been verified by running a t-test with a 0.01 significance level. The null hypothesis is rejected and a p-value of $1.96*10^{-5}$ is returned.

From Figure 2, beginner notes contain more power from the unwanted lower frequencies than the professional standard legato ones. The professional standard legato note samples are much more consistent in the amount of power present associated with the lower frequencies or "playing noise". The mean PSD below the violin's playing range can be used to represent violin timbre and to distinguish between beginner and professional standard legato notes in the dataset. Both the CQT and Welch's PSD based features rely on information obtained via the FFT. The CQT, as used in this work, has been set with a start frequency of 110 Hz and uses eighth tone spacing. The mean of nine specific frequency bins with specific frequency centres below 190 Hz serve well at grouping the dataset's samples according to player type. Using the PSD below 190 Hz to represent the data shows that the beginner note samples in the dataset tend to contain more power below 190 Hz. The CQT frequency bin means indicate less frequency content present in the beginner than in the professional standard legato note samples around the selected centre frequencies. Both measures extract FFT based information from within the same frequency range, but one focuses on power distribution with respect to frequency and the other, on frequency content. From these results, the values obtained from the PSD based feature differentiate effectively between beginner and professional standard legato notes in the dataset.

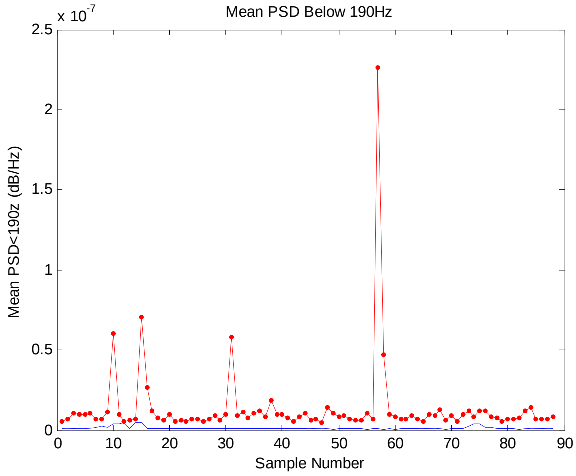

Figure 2: Mean PSD present below 190Hz.

c) SCM Below 190Hz

Jiang *et al.* put forward a filter based spectral contrast measure (SCM) feature in [18]. West *et al.* [19] have also successfully used this feature in the automatic classification tasks of musical signals. As a feature it represents the spectral characteristics of music samples via the relative spectral distribution. It is selected as a violin timbre feature as it has been reported to be designed to give better results than the Mel Frequency Cepstral Coefficients (MFCCs) [19]. It does this by considering the strength of spectral peaks and spectral valleys in each sub-band separately, reflecting the distribution of harmonic and non-harmonic components in a sample. The SCM is considered for its potential as a violin timbre descriptor. The steps involved in extracting this feature are detailed in the afore mentioned papers.

Eight filters are used to divide the frequency domain into sub-bands. The frequency ranges for the filters used are: 0-200 Hz, 200-400 Hz, 400-800 Hz, 800-1600 Hz, 1600-3200 Hz, 3200-6400 Hz, 6400-12800 Hz, and 12800-25600 Hz. The spectral magnitudes of each band are put into descending order according to magnitude. Estimates of the spectral peaks and spectral valleys are obtained using Equation (1) and Equation (2) [18, 19] respectively. In these equations, i is the index, N the window size and α, the neighbourhood factor:

$$p_p = \ln\left(\frac{1}{\alpha N}\sum_{i=1}^{\alpha N} x_{p,i}\right) \qquad (1)$$

$$v_v = \ln\left(\frac{1}{\alpha N}\sum_{i=1}^{\alpha N} x_{v,N-1+i}\right) \tag{2}$$

The inclusion of α, a neighbourhood factor, stabilises the feature by averaging the peaks and valleys within a small region. Jiang *et al.* found that varying α between 0.02 and 0.2 did not influence the performance significantly. In their implementation, α = 0.02 was used. As a starting point in this work, α = 0.02 was taken. Values ranging from α = 0.01 to 0.25 in increments of 0.01 were also used as well as values of α = 0.3 to 0.9 in steps of 0.1. The spectral contrast of each sub-band is given by the difference between the local maxima and minima.

A high SCM reading implies a signal having high peaks, low valleys and strong localised harmonic content. A low SCM reading represents a signal with less harmonic content. Results from all filters have been investigated and the frequency range returning the results of greatest interest is that of 0-200 Hz, as displayed in Figure 3. As in the previous figures, the beginner samples are represented by a dotted line and the professional standard legato ones, by a solid line. The statistical significance of these results has been verified by applying a t-test with a 0.01 significance level. The null hypothesis is rejected and a p-value of $1.67{*}10^{-64}$ has been returned.

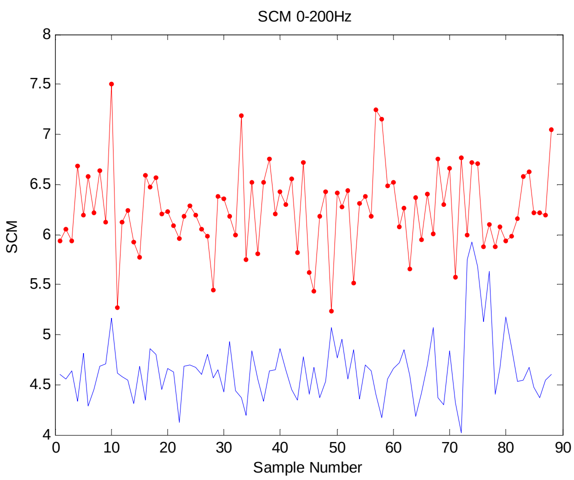

Figure 3: SCM Values for 0-200 Hz filter.

The SCM below 190 Hz reflects similar information to that obtained by the PSD below 190 Hz. The SCM, like the PSD and CQT features, is FFT based and uses both the harmonic and non-harmonic components. Although focusing on the frequency range below 190 Hz, the spectral content is represented as the difference between the harmonic and non-harmonic components, as opposed to the power distribution and harmonic only content in the PSD and CQT based features.

When applying the SCM to the dataset, the most interesting results are returned by the first filter, which is the frequency range below 200 Hz. All the results for this filter from the SCM with α ranging from 0.01 to 0.9 give very good separation between the professional standard legato and the beginner note samples in the dataset. Given that this range includes only the violin's lowest note and below, the content of this region is important.

To investigate this further, a series of filters focusing within this frequency range have been applied. The images displaying the spectral content below 190 Hz, 120 Hz, 85 Hz and 75 Hz as indicated on top of each image are displayed in Figure 4.

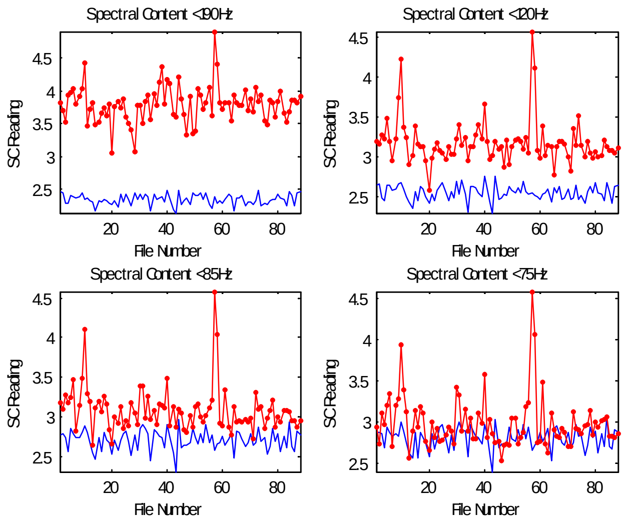

Figure 4: SCM spectral content below 190Hz, 120Hz, 85Hz and 75Hz.

IX CONCLUSIONS

This paper presented a novel approach to investigating the representation of violin timbre as influenced by playing standard. Previous research has focused on the actual playing frequency range of the violin, while this paper focused on the frequency region below the lowest note, G3. Three FFT based features, the CQT, the PSD and the SCM, which have not previously been applied to describe change within the violin's timbre as affected by the player, have been tailored to investigate this. All three features detailed in this paper have shown to accurately distinguish between beginner player note samples and the professional standard legato ones. Therefore, it has been shown that when applying these three features a robust method of categorising can be developed, displaying promising results for use in a computer based violin teaching aid.

REFERENCES

[1] A. Eronen, "Comparison of features for musical instrument recognition", *Proceedings of IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, New Paltz, New York, 2001.

[2] K. D. Martin, Y. E Kim, "Musical instrument identification: a pattern-recognition approach", 136th Meeting ASA, Oct. 1998.

[3] C. M. Hutchins (ed.), V. Benade (ass. Ed.), *Research Papers in Violin Acoustics 1975-1993, Vols. I & II,* Acoustical Society of America, Woodbury NY, 1997.

[4] C. Fritz, I. Cross, B. C. J. Moore, J. Woodhouse, "Perceptual correlates of violin acoustics", *Proceedings of the 9th International Conference of Music Perception and Cognition* (ICMPC), Bologna, 22-26 Aug. 2006

[5] E. V. Jansson "Admittance measurements of twenty-five high quality violins", *Acustica*, 83:337-341, 1997

[6] S. Serafin, M. Burtner, C. Nichols, "Expressive controllers for bowed string physical models", *Proceedings of the 3rd International Conference on Digital Audio Effects*, Limerick, 2001.

[7] R. S. Wilson, "First Steps Towards Violin Performance Extractions Using Genetic Programming" pp. 253-62 in J. R. Koza (ed.) *Genetic Algorithms and Genetic Programming at Stanford 2002*, Stanford, California, 2002.

[8] N.-E. Molin, A. O. Wahlin, E. V Jansson, "Transient wave response of the violin body" (Letters to the Editor), *Journal of the Acoustical Society of America*, 88:5:2479:2481, November 1990.

[9] C. Vergez, P. Tisserand, "The BRASS project, from physical models to virtual instruments: playability", *Lecture Notes in Computer Science*, Vol. 2006 No. 3902, May 2006.

[10] A. Almeida, C. Vergez, R. Causse, X. Rodet, "Physical model of an oboe: comparison with experiments", *International Symposium on Musical Acoustics*, Nara, Japan, April 2004.

[11] N. Giordano, M. Jiang, "Physical modelling of the piano", *European Journal of Applied Signal Processing*, 7:926-933 2004.

[12] G. Agostini, M. Longari, E. Pollastri, "Musical instrument timbres classification with spectral features", EURASIP *Journal on Applied Signal Processing* 2003:1, 1-11.

[13] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland, *The HTK Book (for HTK Version 3.4)*, Dec. 1995, revised Dec. 2006.

[14] B. G. Jackson, J. Berman, K. Sarch, *The ASTA Dictionary of Bowing Terms for Stringed Instruments*, American String Teachers Association, 3rd edition, Tichenor Publishing Group, Bloomington, 1987.

[15] J. Charles, D. Fitzgerald, E Coyle, "Violin timbre space features", ISSC, DIT, Dublin, 2006.

[16] J. C. Brown, "Calculation of a constant Q spectral transform", *Journal of the Acoustical Society of America*, 89, pp. 425-434, 1991.

[17] K. D. Marshall, "Modal analysis of a violin", *Journal of the Acoustical Society of America*, 77(2), Feb. 1985.

[18] D-N. Jiang, L. Lu, H-J. Zheng, J-H. Tao, L-H. Cai, "Music type classification by spectral contrast feature", *Proceedings IEEE International Conference on Multimedia and Expo*, Vol. 1:113:116, 2002.

[19] K. West, S. Cox, "Features and classifiers for the automatic classification of musical audio signals", *Proceedings of the 5th International Conference on Music Information Retrieval* (ISMIR), 2004.