QUANTIFYING VIOLIN TIMBRE

Jane Charles[1]
Dublin Institute of Technology,
School of Control Systems and Electrical Engineering,
Dublin, Ireland

Derry Fitzgerald
Cork Institute of Technology,
Department of Electronic and Computer Engineering,
Cork, Ireland

Eugene Coyle
Dublin Institute of Technology,
School of Control Systems and Electrical Engineering,
Dublin, Ireland

ABSTRACT

Although much research has been carried out on finding features for instrument recognition systems, little work has focused on the violin's timbre space. The effect on sound quality a player may have and the more general area of quantifying the violin timbre space will be investigated in this paper using signal processing techniques. Suitable features from which a computer can assess the quality of a violinist's playing are considered, in particular, the spectral flatness and spectral contrast measures. The eventual outcome of this work can be applied in various systems including the development of a violin or bowed string instrument teaching aid, in automatic music transcription and information retrieval or classification systems.

Keywords – violin, timbre, spectral flatness measure, spectral contrast measure

1. INTRODUCTION

Within the broad area of automatic audio classification, much work has been done in speech recognition, in discriminating between speech and non-speech sounds, and in instrument and environmental sound identification and synthesis. Although many features have been used successfully in automatic instrument recognition systems [1, 2], little work has focused on an individual instrument's timbre. Fritz's PhD thesis [3] looks at the relationship between vocal tract shape and clarinet sound. Current advances in signal processing and interactive computing have enabled the development of much more sophisticated systems and learning aids. One such example has been developed by Hämäläinen *et al.* [4]. This successful real-time singing aid involves pitch-based control of a game character by the user's voice. Little work has been conducted on characterizing or describing the violin's timbre space let alone exploring the relationship between timbre and playing technique. This paper considers the more general area of quantifying the qualitative and subjective nature of violin playing using signal processing techniques. More specifically, two features: the spectral flatness measure (SFM) and the spectral contrast measure (SCM), are considered, as is their suitability as violin timbre features.

2. EXISTING RESEARCH

Much of the existing research to do with violins has been carried out to in order to better understand and emulate the making of top quality sounding instruments. Many methods have been applied to gain insight into the complex interactions between the various components of stringed instruments. Work is ongoing considering the problem of quantifying perception relating to violin sound quality [5, 6]. However, work exploring the effect a player has on the violin sound produced is limited. Finding features which are suitable for quantifying the violin's timbre space involves exploring the effect of a player on sound quality. Steps towards the development of a violin teaching aid which is based on violin pedagogy, sound analysis, and comparison of beginner and good player recordings was presented in [7]. Many features, although very useful in determining one instrument from another [1, 2], are not appropriate for catching the subtleties or nuances due to playing technique or for use within an individual instrument's timbre space. Results have been obtained clearly showing that it is possible for a computer to differentiate between recordings of a beginner note and a good player legato note played on a violin [8, 9].

3. DATA TEST SET

As no suitable data set was readily available, one had to be made. Much thought was given in creating this data set in terms of what was needed, obtainable and viable. The ideal data set would be a type of violin timbre real sound continuum. Unfortunately, this would be very time consuming, if not near impossible to obtain. The first bow stroke a beginner must learn is called *legato*, which literally means 'tied together' or smoothly connected [10]. Mastering this ensures enough bow control upon which the student can develop other bow strokes, such as *staccato* ('disconnected' [10]). 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 data test set was obtained in a controlled environment, consisting of two same sized groups, one with beginner notes and the other with professional standard 'good' player legato notes. The samples all contain one note and are of varying lengths and pitches. A player will never play two notes exactly the same although they may be perceived by a listener as being the same. A beginner does not have the control necessary to achieve this level of accuracy in playing. Hence, it is more appropriate to use features which do not dependent on ether note length or pitch. The data samples were made in a recording studio using two microphones, one directional, the other, omni directional. The tracks were recorded onto DAT, mixed and saved as monophonic wav files. It should also be noted that the recordings were all made in the same studio, using the same microphones and set up as well as the same violin and bow.

4. FEATURE EXTRACTION

Features can be considered as descriptors and can be used for extracting information pertaining to musical signals. In [2],

---

[1] Email: jane.charles@dit.ie

many features for instrument identification purposes have been determined. Many features, although very useful in determining one instrument from another, are not appropriate for understanding the subtleties due to playing technique within an instrument's timbre space. Pitch related or dependent features are of limited use within the context of bowing. As bowing style can induce perceptually acceptable pitch fluctuations but which will cause inaccuracies to occur in most pitch detection systems. Statistical analysis was carried out from which the mean proved to be the most informative and applicable for building a classifier [8]. In [9], the suitability of six features, obtained in different domains, have been considered within the context of the violin timbre space. They include the constant Q transform (CQT), power spectrum density (PSD) estimates, spectral centroid, spectral flatness measure (SFM), and features obtained through cepstral analysis which are based on cepstral coefficients and cepstral log energy. This paper discusses SFM in greater detail and the spectral contrast measure in relation to violin timbre.

4.1. Spectral Flatness Measure (SFM)

The SFM is calculated from the power distribution via Welch's method and is defined as the PSD's geometric mean divided by its arithmetic mean [11]. The steps taken are shown in the figure below:

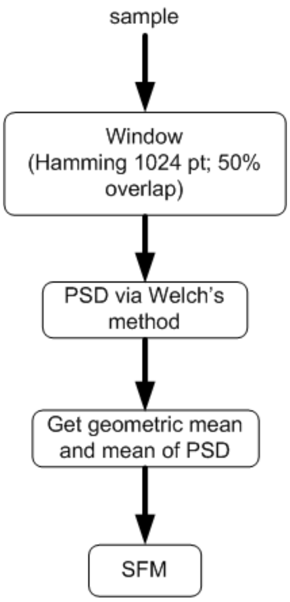

Figure 4.a: Steps taken to obtain spectral flatness readings.

Readings obtained from the SFM indicate how noisy or how close to a pure sinusoid a signal is. As the level approaches 1, the signal is closer to white noise. The closer to zero the reading, the closer the signal is to a pure sinusoid. This has proven to be very useful for separating the transient attack and decay section from the steady-state section in real violin sounds. Figure 4.b below compares a good legato note (top) with a reasonable sounding beginner note (bottom).

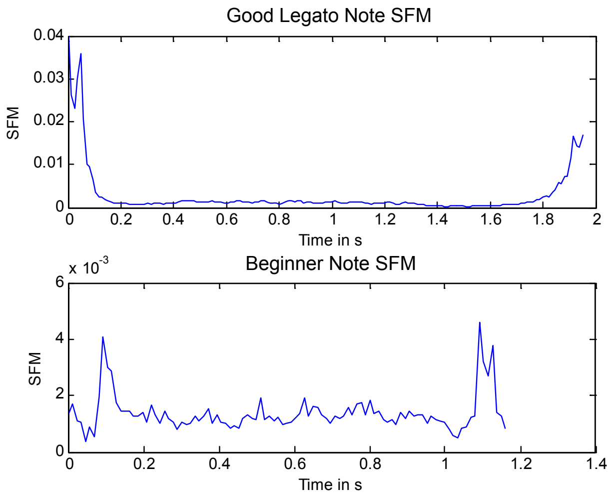

Figure 4.b: A moving SFM for a good legato note (top) and for a reasonable sounding beginner note (bottom).

These images hold much information about the bowing. The steepest changes occur at the beginning and ends of the note and this pattern is repeated throughout the good legato note samples and reasonable sounding beginner samples start approaching this shape too. The starts and ends of notes require more bow control than the middle section. These are also the regions where beginners typically 'crunch' due to lack of bow control. The pressure applied to the string via the bow is not kept the same throughout. The most pressure changes occur when the player in closest to either the tip (top of bow) or towards the heel (bottom of bow) and this is reflected in the SFM readings. The steady-state section of a good legato note, where pressure is applied more consistently, the SFM readings flatten out and approach zero. Attack, steady-state and decay sections become clear in figure 4b, whereas obtaining this information from time or pitch methods for real violin sounds is much more unreliable. This is important in that features can now be applied or developed according to region. For example, more accurate pitch detection can be carried out based only on the steady-state section of the waveform. This is important for string sounds as a significant acceptable fluctuation in pitch does exist due to the attack style. Figure 4.c illustrates the effect of note attack style has on SFM readings. The fast bow stroke takes more time to 'settle down', i.e. go towards a steady-state, than the legato attack. This reflects the force applied to the string and the faster bow stroke causes greater string fluctuations. Should too much pressure be applied to the string, crunching or breaking of sound can be perceived. Figure 4.d shows an example of 'forced' sound on the SFM readings. This 'crunching' example has been deliberately produced by an advanced player. Forcing or crunching causes the SFM reading to increase sharply and remain elevated and unsteady. This image contains several brief sections of crunching.

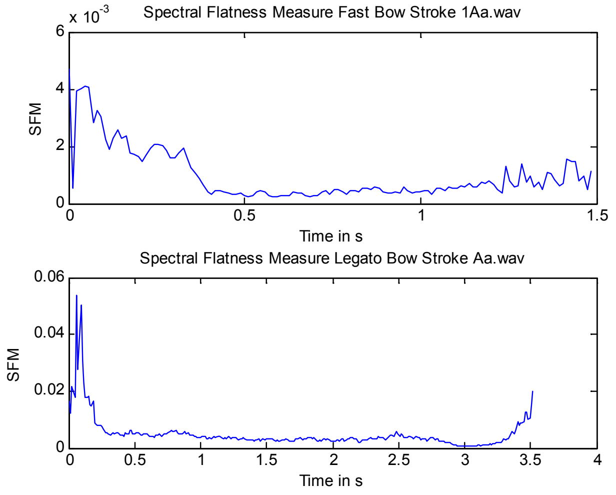

Figure 4.c: the effect of bow stroke style on the SFM reading.

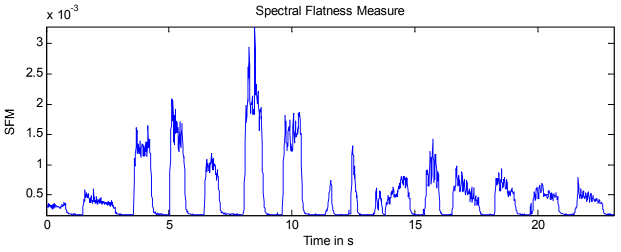

Figure 4.d: SFM readings of 'forced' notes.

Another useful application of the SFM readings is to obtain the note onset for stringed instrument sounds as can be seen in figure 4.e.

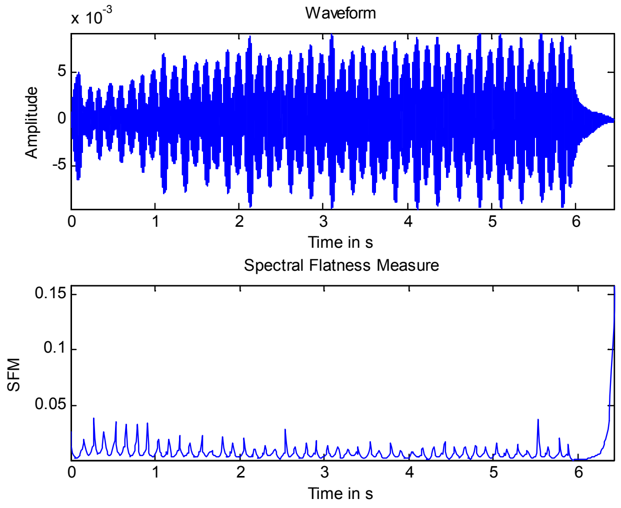

Figure 4.e: Sixteenth notes waveform (top) and its SFM (bottom) showing how the SFM can be used for onset detection for real violin sounds.

Little peaks appear in the SFM readings which line up perfectly with the bow change in figure 4.e. No matter how smooth the bow change, the signal, which prior to this point was going towards sinusoidal wave, hence its SFM reading was approaching zero, is altered. The change and associated disturbance to the steady-state is reflected by a peak or sharp increase in the SFM reading.

4.2. Spectral Contrast Measure (SCM)

Jiang *et al.* put forward an octave filter based spectral contrast feature in [12]. West *et al.* [13] also have successfully used this feature in conjunction with others in the automatic classification tasks of musical signals. It has been selected as a feature as it has been reported to be designed to give better results than the MFCCs approach [13]. The steps involved in extracting this feature are clearly explained in these papers.

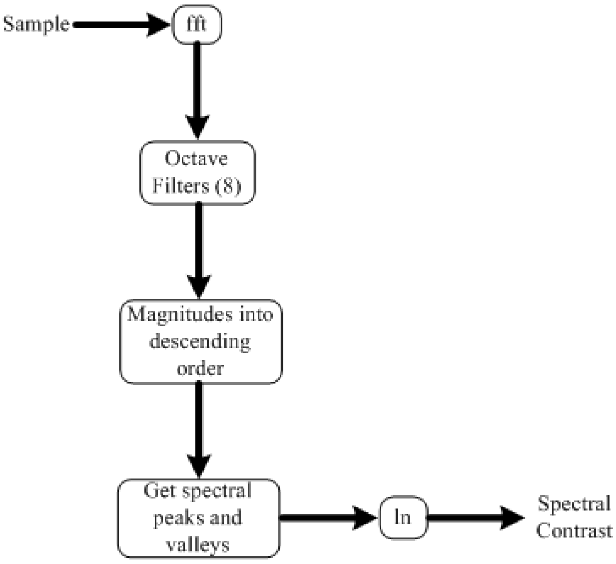

Figure 4.f: steps taken to obtain SCM.

Although both Jiang *et al.* and West *et al.* use a KLT and DCT respectively after obtaining the difference between the peaks and valleys for reducing the covariance between elements of the spectral contrast feature vector, this was not necessary for the data samples used here, as they are all monophonic. The steps applied to obtain the spectral contrast measures of the violin samples can be seen in figure 4.f. The sampling rate of all the data is 44100Hz and eight octave filters have been used to divide the frequency domain into sub-bands. The frequency ranges for the filters are: 0-200Hz, 200-400Hz, 400-800Hz, 800-1600Hz, 1600-3200Hz, 3200-6400Hz, 6400-12800Hz, 12800-25600Hz. The spectral magnitudes of each band are put into descending order to facilitate peak and valley information extraction. The equations applied to obtain the peaks and valleys are shown below in equations (1) and (2) and can be found in [12, 13].

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

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

Where i = index, N = window size and α is a 'neighbourhood factor' and its inclusion stabilizes 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, $\alpha = 0.02$

was applied. The spectral contrast of each sub-band is given by the difference between the peaks and valleys.

$$SpectralContrast_{sub-band} = Peaks_{sub-band} - Valleys_{sub-band} \quad (3)$$

The most interesting results were obtained by the first octave filter (<200Hz) when applying the spectral contrast feature. This filter focused on the frequency content below 200Hz. The lowest note on a violin tuned to A440 is the open G string which is associated with a frequency reading of 196Hz. All the images for this filter from the spectral contrast measure with α ranging from 0.01 to 0.9 give very good separation between the legato good notes and the beginner notes. Given that this range includes only the violin's lowest note and below, this spectral content is of interest. To investigate this further, a series of filters focusing more within this frequency range, were applied from which the spectral contrast feature readings were obtained. A value of $\alpha = 0.2$ worked the best in terms of separating the two groups for the most filters. The filters applied were <190Hz, <160Hz, <120Hz, <90Hz, <85Hz, <75Hz, <60Hz. Excellent separation between the two sample groups was possible until 90Hz at which point the groups are much closer together but the pattern is still discernable. The filters below this showed overlapping frequency content levels. However, on visual inspection a certain amount of similarity in overall shape can be seen in the figures, in particular for the good notes implying a certain consistency in the lower frequency content. This has led to the idea of possibly outlining acceptable 'noise' levels in violin playing. From this it is hoped that playing faults such as 'crunching' can be quantitatively explained.

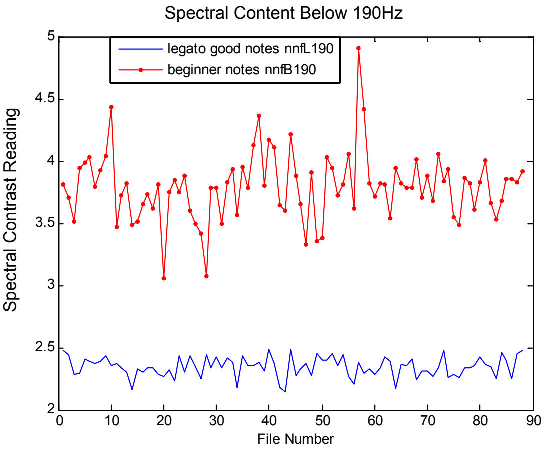

Figure 4.g: Spectral contrast feature using a filter <190Hz, $\alpha = 0.2$.

5. CONCLUSION

The spectral flatness measure and the spectral contrast measure have been shown to be efficient and useful features in relation to real violin sounds. From applying these two features, further segment specific features for real violin sound can be found via the SFM and a possible method of quantifying unwanted noise or crunching in the signal can be obtained via the SCM. However, the violin timbre space remains far from being defined in quantitative terms.

Applications for this work and its continuation include the development of a violin teaching aid, as proposed in [7], for the use in information retrieval and classification systems and for the automatic notation of music or transcription systems.

6. ACKNOWLEDGEMENTS

Special thanks to Ailleen Kelleher, Lioba Petrie, Karla Charles, Graìnne Hope and sound engineer, Owen Tighe.

7. REFERENCES

[1] Eronen A., Klapuri, A. 'Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features', Signal Processing Lab, Tampere University of Technology, Tampere.

[2] Martin, K. D., Kim, Y. E., 'Musical instrument identification: A Pattern-Recognition Approach', 136th Meeting ASA, Oct. 1998.

[3] Fritz, C. 'La clarinette et le clarinettiste: influence du conduit vocal sur la production du son', PhD thesis, L'université de Paris 6 and the University of New South Wales, Dec. 2004.

[4] Hämäläinen, P., Mäki-Patola,T., Pulkki, V., Airas, M. 'Musical Computer Games Played by Singing', Proc. 7th Int. Conf. on Digital Audio Effects (DAFx'04), Naples, Oct. 5-8, 2004.

[5]Woodhouse, J., Cross, I., Moore, B.C.J., Fritz, C. 'Perceptual Tests With Virtual Violins', Proceedings of the Institute of Acoustics, Vol. 28, Pt. 1, 2006.

[6]Fritz, C. *et al.* 'Perceptual Correlates of Violin Acoustics', 9th ICMPC, Bologna, 22-26 Aug. 2006 (pre-publication copy).

[7] Charles, J. A., *et al.* 'Towards a Computer Assisted Violin Teaching Aid', International Symposium on Psychology and Music Education, Nov. 29-30, 2004, Padua, Italy.

[8] Charles, J. A. *et al.* 'Development of a Computer Based Teaching Aid, ViTool', AES 118th Convention, Barcelona, May 28-31, 2005.

[9]Charles, J. A., *et al.* 'Violin Timbre Space Features', ISSC'06, Dublin Institute od Technology, Dublin, June 28-30, 2006.

[10] Jackson, B. G., Berman, J., Sarch, K. *The A.S.T.A. Dictionary of Bowing Terms for String Instruments*, American String Teachers Association, 3rd edition, Tichenor Publishing Group, Bloomington, Indiana, 1987.

[11] Jayant, N. S., Noll, P. *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs NJ, 1984.

[12]Jiang, D-N, Lu, L., Zheng, H-J, Tao, J-H., Cai, L-H. 'Music Type Classification by spectral Contrast Feature', Proceedings IEEE ICME, Vol. 1:113:116, 2002.

[13]West K., Cox, S. 'Features and Classifiers for the automatic Classification of Musical Audio Signals', ISMIR 2004