Towards Prediction of Violin Timbre from Vibrational Measurements

Massimiliano Zanoni, Fabio Antonacci, Augusto Sarti

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano - Piazza Leonardo da Vinci 32 - 20133 Milano, Italy

massimiliano.zanoni@polimi.it

ABSTRACT

This contribution investigates on the acoustics of violin, and more specifically on the relationship existing between vibrational impulse responses and the timbre of the instrument. With respect to previous publications on this topic, we tackle the problem using a feature-based approach. More specifically, we aim at finding the correlation between the features extracted from accelerometric measurements of the bridge mobility and from audio recordings of a prescribed set of performances. Results demonstrate that features describing to the global shape of the spectrum are strongly related. On these descriptors we also show the possibility of predicting the features of audio recordings from the vibrational ones. Experimental data are based on a set of 25 modern violins.

1. INTRODUCTION

The perceived quality of violins is mainly determined by its timbre. Many studies have been proposed in the literature[1, 2]. In this paper we investigate on the relationship existing between vibrational data and audio recordings of performances of violins. The Rayleigh integral [3] suggests that the emitted sound is related to the vibrational velocity field on top and back plates through an integral equation, which, however, cannot be used to predict deterministically the timbre from the vibrational analysis, due to the complexity of the physical and vibrational behavior of the instrument. Instead, works in the literature aim at finding a relationship between timbre, perception and vibrational response of the instrument using tools from machine learning. In an early study in [4] the author demonstrated that peaks in the bridge admittance between 1 kHz and 3 kHz impact on the perception of the timbre. In [5] the author observes that, despite the fact that the bridge response is characterized by relevant peaks and troughs, the perceived quality of the instrument does not change much from note to note played by the musician, suggesting that the timbral quality of the instrument must be attributed to the global shape of the spectrum rather than to its details. In [6] authors relate vibrational and acoustics data of violins by comparing the harmonics of the long-term spectrum of glissando performances with the bridge mobility transfer function, which interestingly turn out to exhibit similar features.

The characterization of the timbre, which is mostly subjective, using objective descriptors extracted from the sound, has gained interest in the research community [7], and is spreading also in Musical Acoustics [2]. In this work we adopt a feature-based approach in order to investigate on the relation between vibrational and audio recordings. As far as vibrational analysis is concerned, we measured the bridge mobility function [8]. Audio data, instead, consist in the recording of excerpts of songs, open strings, and musical scales. Many low-level timbral features have been extracted from both sound recordings and vibrational impulse responses. The Pearson correlation is then analyzed to assess which descriptors of the sound recordings can be predicted from the vibrational data. Results demonstrate that descriptors related to the overall shape of the spectrum exhibit a good degree of correlation, thus confirming the results in [5] and [6]. On these descriptors, we also implemented a curve fitting algorithm, and a good fitting accuracy is found. We believe that this paper sheds light on the possibility for violin-makers to fine tune some parameters of the timbre of the instrument at the different stages of the making process (e.g. additional thinning of plate).

2. DATA ACQUISITION

Our study is based on 25 violins: 16 from the "Triennale Violin-making Competition" held in Cremona, Italy, in September 2015 and 9 from the collection of the school of lutherie.

Vibrational acquisition: in order to conduct the vibration analysis we acquired the bridge mobility, or bridge admittance, of each violin. The instrument was suspended with nylon cables to have a free boundary condition measurement and decoupling the flexible vibration modes from the system rigid motion [9]. The bridge bass side was excited using a small impact hammer (5 g mass) and the response at the other side was captured with a piezoelectric uni-axial accelerometer (0.5 g mass). At least five hammer hits were averaged for each test to reduce the measurement noise [9]. A detail of the excitation and measurement positions is shown in Fig. 1, where also the little foam stripes used to damp the strings on both sides of the bridge are visible.

The H_1 estimator was adopted to derive the Frequency Response Function (FRF) data:

$$H_{oi}(\omega) = \frac{G_{oi}(\omega)}{G_{ii}(\omega)}, \tag{1}$$

where $G_{oi}(\omega)$ is the crossspectral density function between the input driving force i and the output measured vibration o, whereas $G_{ii}(\omega)$ is the autospectral density function of the input force. Fig. 2 shows, as an example, the bridge mobility of three tested violins. The coherence function γ, calculated as

$$\gamma(\omega) = \frac{|G_{oi}(\omega)|^2}{G_{oo}(\omega)G_{ii}(\omega)}, \tag{2}$$

Figure 1. Experimental set-up.

indicates the degree of correlation between input and output and defines the frequency range over which the FRF data can be considered reliable. For the presented data, the upper limit was set to be 2 kHz, whereas the sampling frequency was set to 25.6 kHz.

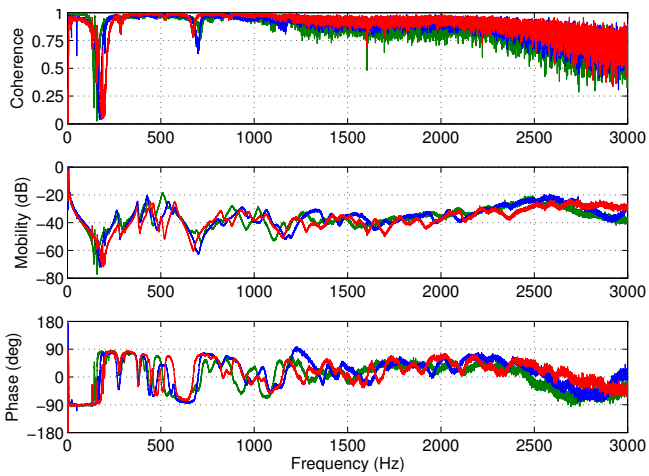

Figure 2. Examples of experimental FRFs for three tested violins; from the top: coherence function, mobility magnitude and phase.

Since some descriptors are extracted from the time domain representation of the signal, we also measured the impulse response $h_{oi}(t)$ as $h_{oi}(t) = \mathcal{F}^{-1}\{H_{oi}(\omega)\}$, where the symbol $\mathcal{F}^{-1}$ denotes the inverse Discrete Fourier Transform.

Audio recordings: instruments were played by a unique professional musician in an acoustic dry room, in order to be independent on the acoustic environment. The musician was asked to play the same pieces of music for all the instruments. The pieces have been properly selected in order to best emphasize the different aspects of the sound of violins: open strings (G,D,A,E), major scale, six musical pieces in different styles (*deciso*, *arpeggiato*, *legato*, *sforzato*, *pizzicato*, *appasionato*, *etereo*). The notes were played at nearly constant intensity, with the bow moving slowly both upward and downward, every time waiting for the sound to completely release before moving on to the next note. For the recordings we used the Beyerdymanic MM1 omnidirectional microphone placed at about 40cm from the instrument. Audio was recorded with an Aphex 188 preamplifier and an Apogee 32 A/D converter at a sampling frequency of 48 kHz and at a bit rate of 24 bits.

3. FEATURE-BASED CHARACTERIZATION

For each instrument, we extracted the set of audio descriptors from audio recordings and vibrational data using a windowing technique. In order to be coherent with the frequency range where the FRF is reliable (see coherence in Fig. 2) we low-pass filtered audio recordings with a cutoff frequency of 2 kHz.

In this study we extend the feature set proposed in [7] by including other descriptors. More specifically, the additional features have been found to be effective in music [10] and sound analysis [11, 2]. A complete description of the additional descriptors is in [10, 12, 13]. The total number of features resulted to be 40. In the following we shortly describe each category.

Noisiness measure: In order to provide a measure of the noisiness of sound we use Zero Crossing Rate (ZCR). In addition, we compute the Spectral Flatness and Spectral Crest to measure the similarity between the magnitude of the spectrum under analysis and a flat one. **Statistical moments:** in order to compactly represent the shape of the spectrum, we compute Spectral Centroid, Spectral Spread, Spectral Average Deviation, Spectral Skewness, and Spectral Kurtosis. **Spectral shape:** besides the statistical moments, also other descriptors provide a compact representation of the shape of the spectrum. We include Spectral Rolloff, Spectral Smoothness, Spectral Slope, Spectral Irregularity K [13] and Spectral Irregularity J [14]. **Summary of the spectrum:** in order to provide a generic summary of the spectral components, MFCCs (20 coefficients) and Spectral Contrast (7 descriptors) are also considered in this study. MFCCs are based on Mel-Frequency scale, based on the auditory model. They are obtained as the coefficients of the discrete cosine transform (DCT) applied to a reduced Power Spectrum, which is derived as the log-energy of the spectrum. Spectral Contrast [12] is defined as the dynamics of spectral peaks and valleys separated into different frequency sub-bands of interest. Samples from each sub-band are sorted in descending order and peaks and valleys of the k-th sub-band are computed as

$$\mathrm{P}_k = \log\left\{\frac{1}{\alpha N_k}\sum_{j=1}^{\alpha N_k}|S_{k,j}|\right\}, \tag{3}$$

$$\mathrm{V}_k = \log\left\{\frac{1}{\alpha N_k}\sum_{j=1}^{\alpha N_k}|S_{k,N_k-j+1}|\right\}. \tag{4}$$

Finally, the Spectral Contrast is computed as

$$\mathrm{SC}_k = \mathrm{P}_k - \mathrm{V}_k, \tag{5}$$

where α is a corrective factor used in order to ensure the steadiness of the feature [12], $S_{k,j}$ is the j-th sample of the DFT within the k-th sub-band and N_k is total number of samples in the k-th sub-band.

4. RESULTS

Correlation analysis: fig. 3 shows the Pearson correlation index between vibration- and audio-based features. As for

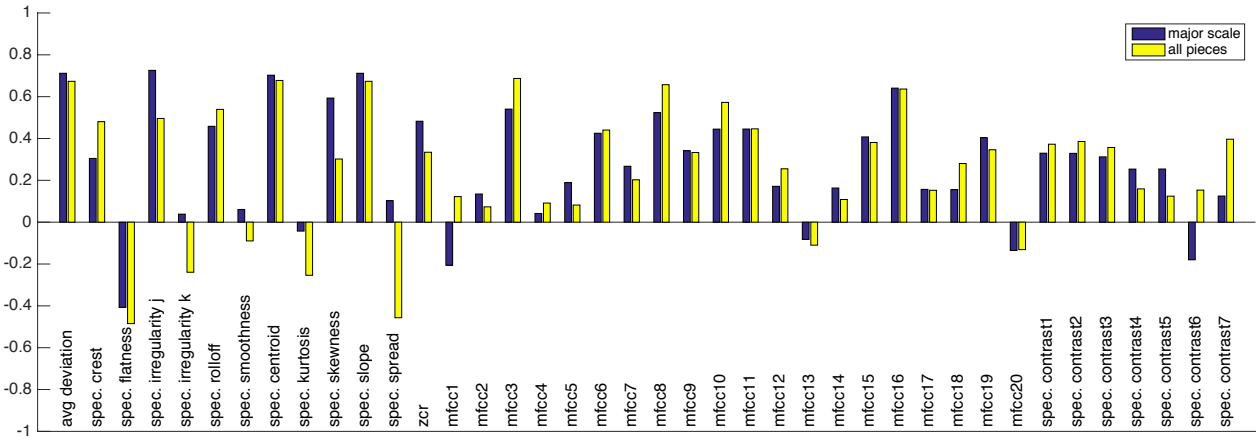

Figure 3. Pearson Correlation Index on the major scale and on all pieces performances.

the audio features, we performed the analysis for each of the performances described in Section 2. However, for reasons of space, only the correlation of descriptors extracted from the major scale and from all performances are shown.

Among the 40 features, several show a good correlation index. These are Average Deviation, Spectral Irregularity J, Spectral Centroid, Spectral Slope, and third, eighth and sixteenth coefficients of MFCC. It is worth noticing that these features convey global information about the spectrum, thus confirming the results in [5] and [6]. It is worth noticing that, despite a slight decrease, the set of most relevant descriptors remains unaltered when the analysis is extended to all the music pieces. **Curve Fitting Analysis**: in order to provide further insight on the correlation analysis, we also tested the accuracy of different curve fitting models on the features listed above. The adopted fitting models are linear, quadratic, cubic, and quartic polynomials and the exponential. In preliminary results we noticed that the use of higher-order polynomials does not bring relevant advantages in terms of fitting accuracy, thus we limited our analysis to the fourth order. Data have been previously normalized between 0 and 1. In order to be more robust to the presence of outliers in the dataset, we performed outlier identification based on the residuals of the fitting curves. Let $O = \{(x_i, y_i) \colon 1 \leq i \leq N\}$ be the dataset, being N the cardinality, and y and x the audio- and vibration-based features. Let $f(x)$ be the fitting curve and $f(x_i)$ be the corresponding point in the model for x_i, the residual r_i is computed as $r_i = x_i - f(x_i)$. Points in the distribution are considered outliers if $|r_i| > \mu_{|r_i|} + 1.5\sigma_{|r_i|}$, where $\mu_{|r_i|}$ and $\sigma_{|r_i|}$ are the mean and the standard deviation of the absolute values of r_i respectively.

In order to evaluate the quality of the fitting we have adopted two metrics, which are the ordinary Coefficient of Determination (R^2) and the Root Mean Square Error ($RMSE$), given by

$$R^2 = 1 - \frac{\sum_{i=1}^{N} |y_i - f(x_i)|^2}{\sum_{i=1}^{N} |y_i - \overline{y}|^2}, \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} |y_i - f(x_i)|^2}{N}}, \tag{7}$$

where $\overline{y}$ is the mean value of the considered vibration-based descriptor. R^2 ranges from 0 (fitting is not possible) to 1 (data can be fitted without any error), while $RMSE$ has the lower bound of 0, meaning a perfect fitting. We conducted the analysis both removing and keeping outliers. For reasons of space, we present here only the fitting with features extracted on the major scale performance. Fig. 4 shows the R^2 for the seven selected descriptors and for the different fitting models. The best accuracy is observed for the quartic curve. This is true for all the descriptors, but especially for MFCC 8 and MFCC 3. We repeated the experiment when data that have been

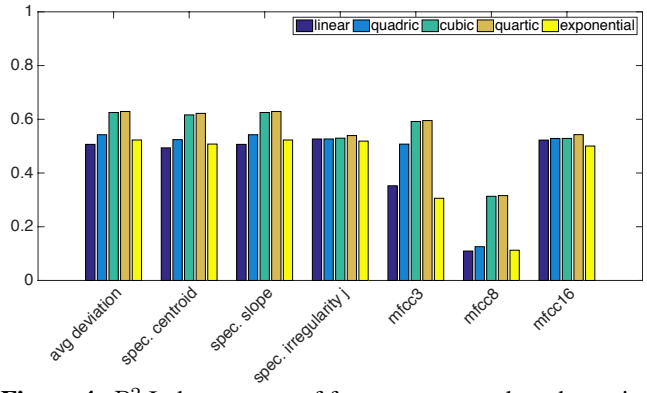

Figure 4. R^2 Index on a set of features extracted on the major scale performance.

recognized as outliers have been removed from the dataset. Results are reported in fig. 5. It is possible to observe a general increase of the R^2 metric.

Finally, Fig. 6 and Fig. 7 show the Root Mean Square Error with and without outliers. Even if the two metrics are not completely independent (see eq. (6) and (7)), it is worth noting that the RMSE and the R^2 metrics suggest that the best prediction is achievable for the MFCC16 and Average Deviation, respectively.

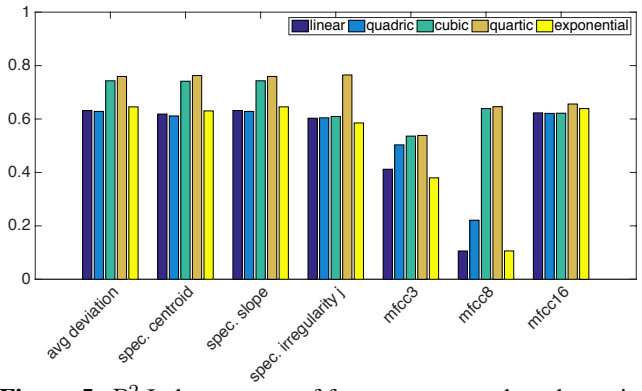

Figure 5. R^2 Index on a set of features extracted on the major scale performance without outliers.

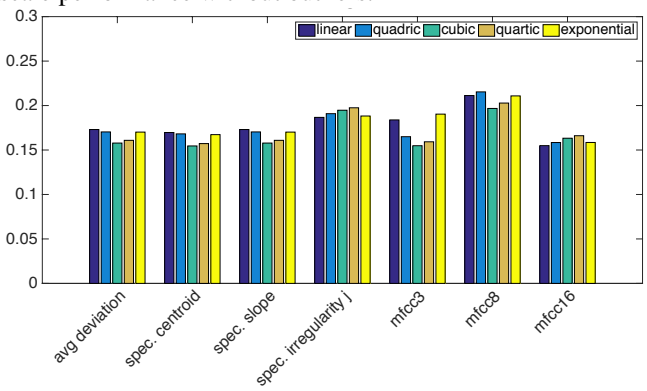

Figure 6. Root Mean Square Error (RMSE) on a set of features extracted on the major scale performance.

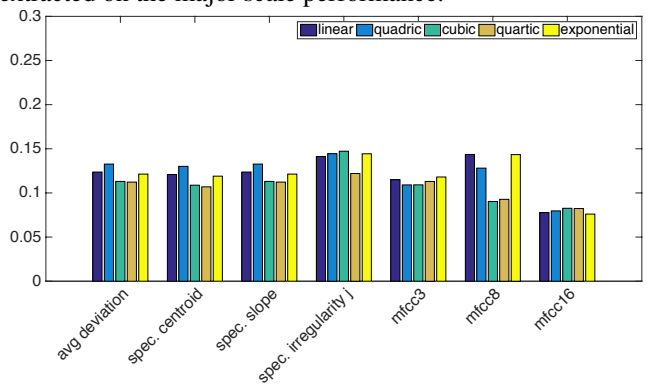

Figure 7. Root Mean Square Error (RMSE) on a set of features extracted on the major scale performance without outliers.

5. CONCLUSIONS AND FUTURE WORKS

In this contribution we investigated on the relationship existing between vibroacoustics and audio recordings of violins. Using a feature-based approach we demonstrated that the global shape of vibrational and audio spectra exhibit a good degree of correlation. This result paves the way to the possibility of predicting the timbre of the instrument also from accelerometric measurements. Some results in this direction are included in this paper.

REFERENCES

[1] J. Štěpánek, Z. Otčenášek, and A. Melka, "Comparison of five perceptual timbre spaces of violin tones of different pitches," *The Journal of the Acoustical Society of America*, vol. 105, no. 2, pp. 1330–1330, 1999.

[2] M. Zanoni, F. Setragno, F. Antonacci, A. Sarti, G. Fazekas, and M. Sandland, "Training-based semantic descriptors modeling for violin quality sound characterization," in *in Procideengs of the Audio Engineering Society Convention (AES)*, Warsaw, Poland, 2015.

[3] E. G. Williams, *Fourier Acoustics: sound radiation and nearfield acoustical holography*. Academic Press, 1999.

[4] M. Hacklinger, "Violin timbre and bridge frequency response," *Acta Acustica united with Acustica*, vol. 39, no. 5, pp. 323–330, 1978.

[5] C. Gough, "The violin: Chladni patterns, plates, shells and sounds," *The European Physical Journal Special Topics*, vol. 145, no. 1, pp. 77–101, 2007.

[6] B. Elie, F. Gautier, and B. David, "Acoustic signature of violins based on bridge transfer mobility measurements," *The Journal of the Acoustical Society of America*, vol. 136, no. 3, pp. 1385–1393, 2014.

[7] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 16 November 2011.

[8] J. Woodhouse, "The acoustics of the violin: a review," *Reports on Progress in Physics*, vol. 77, no. 11, 2014.

[9] R. Corradi, A. Liberatore, S. Miccoli, F. Antonacci, A. Canclini, A. Sarti, and M. Zanoni, "A multidisciplinary approach to the characterization of bowed string instruments: the musical acoustics lab in cremona," in *Proceedings of the 22nd International Congress on Sound and Vibration (ICSV22)*, 2015, pp. 2633–2640.

[10] H.-G. Kim, N. Moreau, and T. Sikora, *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. John Wiley & Sons, 2005.

[11] Y. Kong, A. Mullangi, J. Marozeau, and M. Epstein, "Temporal and spectral cues for musical timbre perception in electric hearing," *Journal of Speech Language and Hearing*, vol. 54, no. 3, Jun 2011.

[12] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, , and L. H. C. in Proc. of IEEE International Conference Multimedia Expo, "Music type classification by spectral contrast features," in *Proc. of IEEE Internationa Conference Multimedia Expo*, 2002, p. 113116.

[13] J. Krimphoff, S. McAdams, and S. Winsberg, "Caracterisation du timbre des sons complexes. ii analyses acoustiques et quantification psychophysique." *Journal de Physique IV, Colloque C5*, vol. 4, 1994.

[14] K. Jensen, "Timbre models of musical sounds," Ph.D. dissertation, University of Copenhagen, 1999.