

TRANSCRIPTION AND EXPRESSIVENESS DETECTION SYSTEM FOR VIOLIN MUSIC

Isabel Barbancho, Cristina de la Bandera, Ana M. Barbancho, Lorenzo J. Tardón

E.T.S.I. Telecomunicación, Dpt. Ingeniería de Comunicaciones,
Universidad de Málaga, Campus de Teatinos s/n, 29071 Málaga, Spain

ABSTRACT

In this paper, a transcription system for music played by violin is presented. The transcription system not only detects the pitch and duration of the notes but also identifies successfully the employed technique to play each note: *détaché* with and without accent and with and without vibrato, *pizzicato*, *tremolo*, *spiccato* and *flageolet-töne*. The transcription system is based on a combined analysis of the time domain and frequency domain properties of the music signal.

Index Terms— Music, Acoustic signal processing, Modeling, Audio systems, Signal classification

1. INTRODUCTION

The violin is the most important of the stringed instruments and it is, after the piano, the most popular instrument for music learners [1]. In this paper, an expressiveness transcription system for the violin is presented. The proposed transcription system is able to detect not only the played pitch, as other transcription systems [2], but also the technique employed in the performance.

In section 2, the time and frequency tools that constitute the basis of the detection system are explained. In section 3, the different violin performances are parametrized. In section 4, the expressiveness transcription algorithm is presented and section 5 will describe the performance of the proposed system. Finally, section 6 summarizes the main conclusions.

2. TIME AND FREQUENCY TOOLS FOR THE CHARACTERIZATION OF VIOLIN NOTES

In this section, the time and frequency tools employed to characterize each of the playing techniques are described.

2.1. Time domain characterization

The duration and the shape of the envelope of the violin notes contains information about the technique employed to play the note under analysis. The envelope model proposed by Jensen [3] will be considered. To obtain the envelope, a Butterworth filter of order 5 and cutoff frequency $11Hz$ has been used. Before the filter, the audio samples are squared. After the filter, the signal is normalized so that the maximum is 1 and the samples under a 2.5% of the maximum amplitude are

eliminated. The attack time (T_a) is considered from the first sample that reaches a 10% of the maximum amplitude until the first sample that reaches a 85% of the maximum amplitude. The release time (T_r) is considered from the last sample that reaches a 70% of the maximum amplitude to the last one over the 10% of the maximum amplitude. The time between T_a and T_r is the sustain time (T_s). The times T_a , T_r and the shape of the envelope during T_s will depend on the bowing.

2.2. Frequency domain characterization

The FFT of the violin signal is obtained, then the magnitude is calculated and it is warped as proposed in [4]. The frequency axis is converted into MIDI numbers according to eq. (1). Taking into account that the frequency range of the violin is G3-C8, the MIDI numbers of interest range between 55 and 108 according to:

$$MIDI = 69 + 12\log_2(f/440) \quad (1)$$

The spectrum is simplified on the basis of the MIDI number as follows. For a certain MIDI number n_{MIDI} , the spectrum between $n_{MIDI} - 0.5$ and $n_{MIDI} + 0.5$ is considered and the maximum spectrum value of that interval is assigned to n_{MIDI} . The MIDI number spectrum (or MIDI simplified spectrum) of each note will have 54 samples. Let $S_{MIDI}(n)$ denote the MIDI simplified spectrum of a given audio sample with n standing for the MIDI number used to represent each MIDI band on this representation. From this simplified spectrum, the pitch and the spectrum width around the pitch will be calculated.

2.2.1. Pitch extraction

In order to detect the pitch of the played note, first a threshold $pitch_{th} = 0.7\max\{S_{MIDI}(n)\}$ is set and the samples over this threshold are determined. Then, an iterative process starts: the largest sample over the threshold is localized and stored and the four adjacent samples are eliminated. Then the largest sample from the remaining is localized and the same procedure is followed. The MIDI numbers of the stored samples are sorted in ascending order and the distances in MIDI number between them are calculated.

Let f_{pitch} denote the fundamental frequency of the played note, the expected harmonics for a violin will be f_{pitch} , $2f_{pitch}$, $3f_{pitch}$ and $4f_{pitch}$. Let n_1 , n_2 , n_3 and n_4 stand for the MIDI numbers that corresponds to those frequencies,

This work was supported by the Ministerio de Educación y Ciencia of the Spanish Government under Project No. TSI2007-61181.

the expected distances between them are $d_{\{n_1, n_2\}} = 12$, $d_{\{n_2, n_3\}} = 7$, and $d_{\{n_3, n_4\}} = 5$. In most of cases, the smallest MIDI number corresponds to the pitch of the played note, so $n_{pitch} = n_1$. However, not always the first harmonic is detected since sometimes it is not the strongest one or even it has not enough power to exceed the threshold. This happens mainly in the low frequency range. So, the distances $d_{\{n_i, n_{i+1}\}}$ between the detected harmonics MIDI number n_i are calculated and if the MIDI number difference between the first and the second harmonic is $d_{\{n_1, n_2\}} = 7$, then it is assumed that the fundamental frequency has been not detected and the pitch is determined as $n_{pitch} = n_1 - 12$.

2.2.2. Spectral width calculation

This parameter, that will be noted as M_{BW} , measures the number of consecutive samples around each maximum in the simplified spectrum over a threshold defined as $M_{BW,th} = 0.25 \max\{S_{MIDI}(n)\}$, up to a maximum of 5 samples on each side. The detection system considers only the bandwidth around the largest peak in the simplified spectrum.

3. PLAYING TECHNIQUE CHARACTERIZATION

In this section, we will characterize the different playing techniques taking into account their characteristics in time and frequency domains explained in section 2. This characterization will be the basis of the detection technique that will be studied in section 4.

3.1. Détaché

In this technique, the bow is in contact with the string without strong accents nor irregularities. That technique produces a uniform sound that will have a smooth envelope and a clean spectrum. Figure 1 presents the envelope and the spectrum of an A4 note played in *détaché*. This note corresponds to the middle range of the notes played with a violin. There are no special characteristics if notes with higher- or lower- pitches are played. In this technique, the time duration of the note is variable. Also, the attack and release times are not determined in this playing technique because they depend on the violinist and on the intention with which the note is played (with accent or without accent). A note played in *détaché* without vibrato presents a very clean spectrum (spectral width of the maximum spectral component small) and the harmonics can be easily identified. Also, in *détaché* the legato playing is very common, i.e., to play without any perceptible interruption between the notes. In this case, in which the violinist changes between notes without stopping the bowing, there is no attack time. Therefore, in order to identify a note played in *détaché*, we will study the other techniques and a note will be determined to be played in *détaché* without vibrato if it does not fulfill the characteristics of the other techniques.

3.2. Pizzicato

In this case, the note is not played with the bow but the string is plucked with the index finger of the right hand.

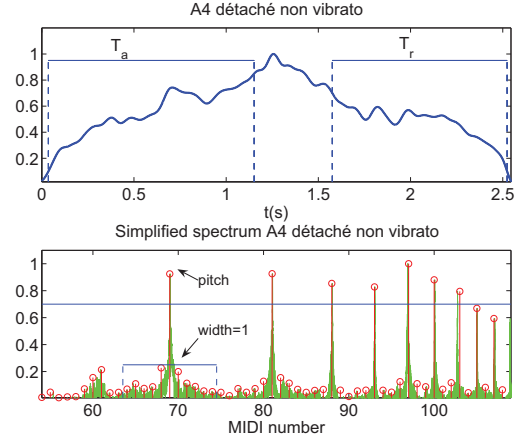


Fig. 1. A4 played in *détaché*

In Fig. 2, the envelope and the spectrum of a C4 played in pizzicato are shown. The envelope characteristics for a note played in pizzicato are the following: the attack time is $0.02s \leq T_a \leq 0.03s$ and the envelope during the attack time is strictly increasing due to the free oscillation of the string. The note duration is short and the release time is $T_r < 0.02s$. About the spectral characteristics, it can be observed that there are many spectral components whose power is over $0.25 \max\{S_{MIDI}(n)\}$. Specifically, $M_{BW} \geq 8$.

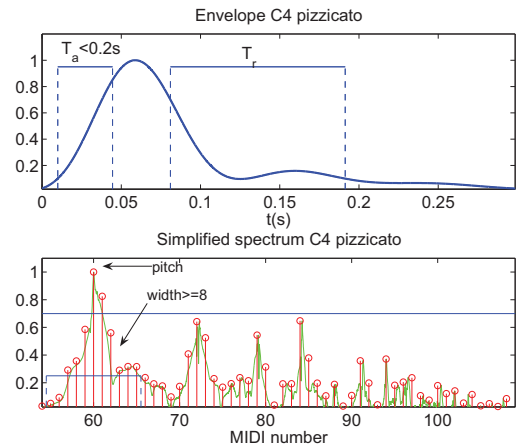


Fig. 2. C4 played in pizzicato

3.3. Tremolo

This technique consists on the quick reiteration of the same tone produced by a rapid up-and-down movement of the bow. In Fig. 3, the envelope and the spectrum of a C4 tremolo played is shown. The envelope of a note played with this technique shows relevant amplitude fluctuations. The attack and release times are short in this technique due to the rapid movement of the bow. Thus the attack time, $0.05s \leq T_a \leq 0.1s$, is the most relevant characteristic in order to characterize this technique. In the simplified spectrum, $2 < M_{BW} < 8$.

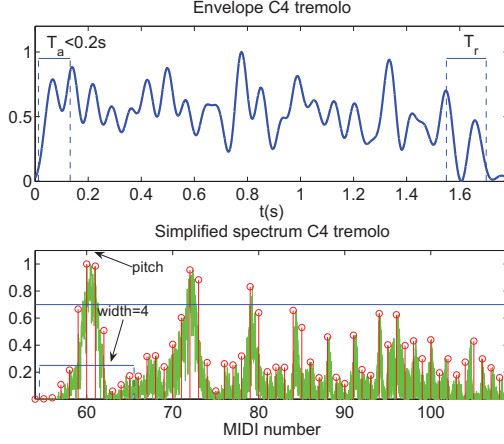


Fig. 3. C4 played in tremolo

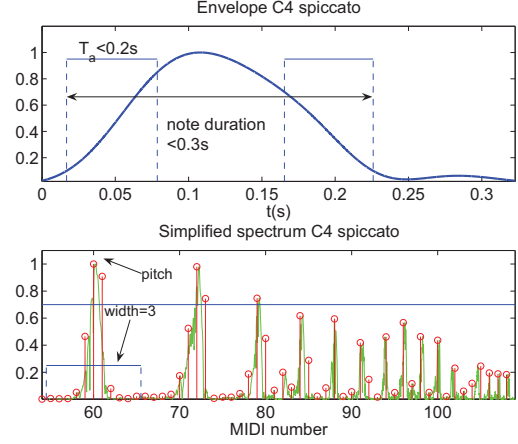


Fig. 4. C4 played in spiccato

3.4. Spiccato

In this technique, the bow is dropped on the string and it is lifted again after each note. Therefore, the first thing to take into account for this technique is the short duration of the note. In Fig. 4, the envelope and the spectrum of a C4 spiccato played is shown. The attack time is $0.04s < T_a < 0.07s$. The envelope is similar to the pizzicato one, but in this case the attack time is a little bit bigger. The whole duration of the note ($T_d = T_a + T_s + T_r$) is $T_d \leq 0.3s$. If the attack is short but $T_d > 0.3s$, the system will consider that the note is played in détaché with accent. As it can be observed in Fig.4, the spectrum in this technique is quite clean and the main frequency component is narrow ($M_{BW} < 8$).

3.5. Flageolet-töne

Flageolet tones are generated when a player gently damps the vibrating string at a certain location so that only those modes that have a nodal point at the damping location remain ringing. In Fig. 5, the envelope and the spectrum of a E6 played in flageolet is shown. The lowest note that can be flageolet played in a violin is G4 (MIDI number 67). It can be observed that the envelope is quite irregular because the intensity of the sound is unstable due to the small contact of the finger and the bow over the string. The duration of the note and the attack time in this technique vary a lot with the musician, so durations are not relevant characteristics for the detection of this technique. About the spectral characteristics, the finger must brush the string just in its node, so the sound of the note will be perfectly tuned and $M_{BW} \leq 2$.

3.6. Vibrato

Vibrato stands for a slight fluctuation of pitch produced on sustained notes by an oscillating motion of the left hand. The detection tool will consider the melodic vibrato that is used in long notes. Therefore, the detection will be focus on vibrato

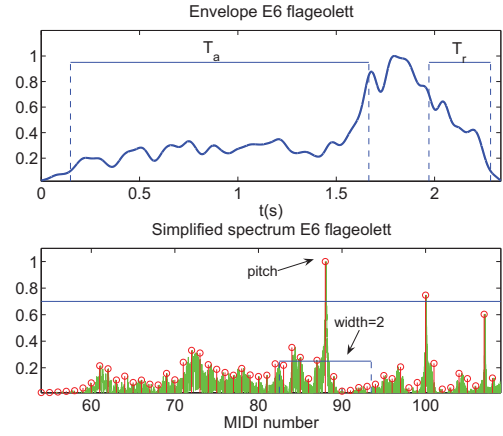


Fig. 5. E6 played in flageolet

employed in notes played in détaché. Fig. 6 shows the envelopes of a C4 and a D7 played with and without vibrato. On not very high-pitch notes, the difference in the envelope when a note is played with vibrato and without vibrato is clear, otherwise, the difference is not so obvious, but with vibrato, the oscillations are more accentuated. Due to this fact, two different conditions on the amplitude of the oscillations are set. Let A_{av} denote the average amplitude of the envelope during T_s , if there is any sample in the envelope during T_s out of the interval $(A_{av} - A_{th}, A_{av} + A_{th})$, with $A_{th} = 0.2$ for notes under A6 (midi number 93) and $A_{th} = 0.3$ for notes above A6, it is decided that the note has been vibrato played. In the spectrum, there are no relevant features to decide about the vibrato.

4. DETECTION TECHNIQUE

In this section, we explain how the characteristics detailed in the previous sections are analyzed in order to determine the technique employed to play the note under analysis. Table 1 summarizes the most relevant characteristics in order to detect the technique. In the table, SA stands for short attack, SD for

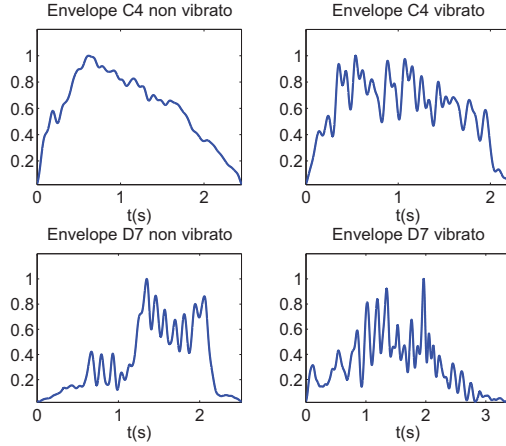


Fig. 6. Notes of different pitch played with and without vibrato

short duration, NBSC for narrow band spectral components, WBS for wide band spectral component and EF for envelope fluctuations. The parameters are obtained as explained previously and following the flowchart shown in Fig. 7.

Charac.	SA	SD	NBSC	WBS	EF
Tech.					
Pizzicato	Yes	Yes	No	Yes	No
Tremolo	Yes	No	No	No	Yes
Spiccato	Yes	Yes	No	No	No
Flageolett	No	No	Yes	No	No
Détaché	No	No	No	No	No
Détaché with accent	Yes	No	No	No	No
Vibrato	No	No	No	No	Yes

Table 1. Main characteristics of the different playing techniques.

5. EVALUATION OF THE SYSTEM

The evaluation of the expressiveness transcriptor described in this paper has been made using five different violins, different musicians and different recording qualities. Professional recorded samples of violins 1 to 3 are taken from the Musical Instrument Sound Data Base RWC-MDB-1-2001-W05 [5]. Violin1 is an italian violin made by J.F. Pressenda, Violin2 is a XVIII century violin made by Carcassi and Violin3 was made in Tokio by Fiumebianca. Home made recording using a microphone RODE NT1 of two other violins have been used: a French violin from the beginning of XX century and a Rapsody electric violin. The system has been tested over the whole violin frequency range and all the explained playing techniques. The overall performance of the system is 95.4%. Table 2 shows the detailed results. The table shows the errors in pitch detection, the error in technique detection and the percentage for wich notes and technique are correctly detected. Observe that more errors are encountered in the expressiveness technique detection than in the pitch detection. This is due to the fact that, depending on the musician and on the violin, the technique characteristics vary and it is difficult to set the conditions to decide the expressiveness technique. Also, as expected, there are more errors in the home made

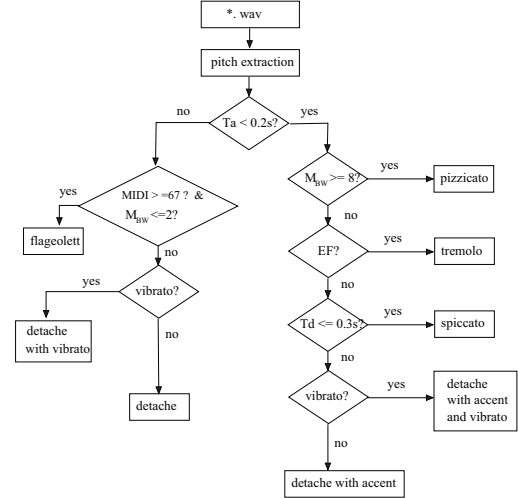


Fig. 7. Expressiveness decision flowchart.

recordings than in the professional ones.

	Recording	Pitch errors	Technique errors	Error free
Violin1	PR	4%	2%	94%
Violin2	PR	0%	0%	100%
Violin3	PR	0%	8%	92%
Violin4	HMR	6%	13%	81%
Violin5	HMR	0%	10%	90%

Table 2. Performance of the detection system. PR stands for professional recordings and HMR for home made recordings.

6. CONCLUSION

In this paper, a tool for the pitch and expressiveness transcription of violin music has been presented. The system employs time and frequency characteristics of the played note to determine both the pitch and the playing technique. The system has a good performance on home made and professional recordings. The proposed system can be extended to the other instruments of the violin family.

7. REFERENCES

- [1] W. Apel, *Harvard Dictionary of Music, Second edition*, The Belknap Press of Harvard University Press, 1972.
- [2] W. J. J. Boo, Y. Wang, and A. Loscos, "A violin music transcriber for personalized learning," *IEEE Int. Conf. on Multimedia and Expo*, pp. 2081 – 2084, July 2006.
- [3] K. Jensen, "Envelope model of isolated musical sounds," *Proc. of DAFx99*, 1999.
- [4] A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. on Speech and Audio Processing*, vol. 11, pp. 804–816, November 2003.
- [5] M. Goto, "Development of the RWC music database," *Proc. of the 18th Int. Congress on Acoustics*, pp. I–553–556, April 2004.