LEGATO AND GLISSANDO IDENTIFICATION IN CLASSICAL GUITAR

Tan Hakan Özaslan and Josep Lluis Arcos
Artificial Intelligence Research Institute, IIIA.
Spanish National Research Council, CSIC.
{tan,arcos}@iiia.csic.es

ABSTRACT

Understanding the gap between a musical score and a real performance of that score is still a challenging problem. To tackle this broad problem, researchers focus on specific instruments and/or musical styles. Hence, our research is focused on the study of classical guitar and aims at designing a system able to model the use of the expressive resources of that instrument. Thus, one of the first goals of our research is to provide a tool able to automatically identify expressive resources in the context of real recordings. In this paper we present some preliminary results on the identification of two classical guitar articulations from a collection of chromatic exercises recorded by a professional guitarist. Specifically, our system combines several state of the art analysis algorithms to distinguish among two similar guitarists' left hand articulations such as legato and glissando. We report some experiments and analyze the results achieved with our approach.

1. INTRODUCTION

An affective communication between listeners and performers can be achieved by the use of instruments' expressive resources [1, 2, 3]. Expressive resources play also an important role to clarify the musical structure of a piece [4, 5, 6]. Although each instrument provides a collection of specific expressive capabilities, its use may vary depending on the musical genre or the performer.

Our research on musical expressivity is focused on the study of classical guitar and aims at designing a system able to model the use of the expressive resources of that instrument. As a first stage of our research we are developing a tool able to automatically identify the use of guitar articulations.

There are several studies on plucked instruments and guitar synthesis such as on extraction of expressive parameters for synthesis [7, 8]. However, expressive articulation analysis from real guitar recordings has not been fully tackled. The analysis of a guitar performance is complex because guitar is an instrument with a rich repertoire of expressive articulations.

In guitar playing both hands are used: one hand is used to press the strings in the fretboard (commonly the left hand) and the other to pluck the strings. Strings can be plucked using a single plectrum called a flatpick or by directly using the tips of the fingers. The hand that presses the frets is mainly determining the notes while the hand that plucks the strings is mainly determining the note onsets and timbral properties. However, left hand is also involved in the creation of a note onset and different expressive articulations like legato, glissando, grace notes, or vibratos [8].

In a previous research [10], we proposed a system able to detect attack-based articulations and distinguish among legato and grace notes. The goal of this paper is to extend the capabilities of the existing system to distinguish among legato and glissando articulations. In both, legato and glissando, left hand is involved in creation of the note onset.

In the case of ascending legato, after plucking the string with the right hand, one of the fingers of the left hand (not already used for pressing one of the frets), presses a fret causing another note onset. Descending legato is performed by plucking the string with a left-hand finger that was previously used to play a note (i.e. pressing a fret).

The case of glissando is similar but this time after plucking one of the strings with the right hand, the left hand finger that is pressing the string is slipped to another fret also generating another note onset. Notice that we are not considering here grace notes that are played in a similar way than glissando.

When playing legato or glissando on guitar, it is common for the performer to play more notes within a beat than the stated timing enriching the music that is played. A powerful legato and glissando can be differentiated between each other easily by ear. However, in a musical phrase context where the legato and glissando are not isolated, it is hard to differentiate among these two expressive articulations.

The structure of the paper is as follows: Section 2 describes our methodology for legato and glissando determination and differentiation. Specifically, our approach uses aperiodicity information to identify articulations, histograms to compute the density of the peak locations, and a symbolic aggregate approximation (SAX) representation to characterize the articulation models. Next, Section 3 focuses on the experiments conducted to evaluate our system. Last section, Section 4, summarizes current results and proposes the next research steps.

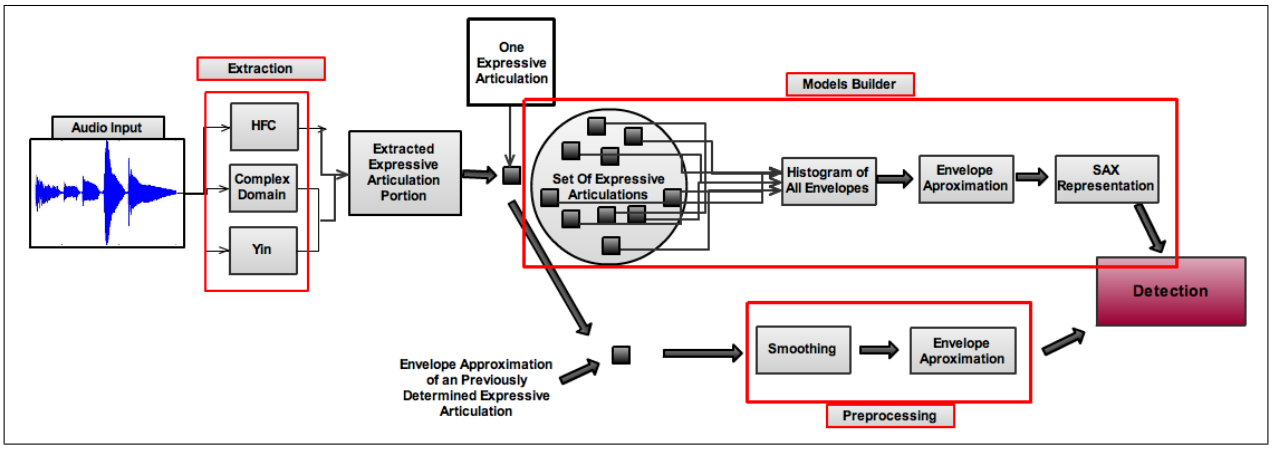


Figure 1: Main diagram of our model, which contains three sub models; Extraction, Model Builder and Preprocessing

2. SYSTEM ARCHITECTURE

In this paper we propose a new system able to identify two expressive articulations: legato and glissando. To that purpose we use a *Region Extraction* module that is part of a previous development [10]. The regions identified by the region extraction module are the inputs to the new components: the *Models Builder* and the *Detection* component (see Figure 1). In this section, first we briefly present the region *Extraction*. Next, we describe our preliminary research to select the appropriate descriptor to analyze the behavior of legato and glissando. Finally, we explain the new two components, Model Builder and Detection.

2.1 Extraction of Candidates

Guitar performers can apply different articulations by using both of their hands. However, the kind of articulations that we are investigating (legato and glissando) are performed by the left hand. Although they (legato and glissando) can cause onsets, these onsets are not as powerful in terms of energy and also have different characteristics in terms of harmonicity, comparing to the plucking onsets [11]. Therefore, we need an onset determination algorithm suitable to differentiate between plucking onsets and left-hand onsets.

The first task of the extraction module is to determine the onsets caused by the plucking hand, i.e. right hand onsets. As right hand onsets are more percussive than left hand onsets we use a measure appropriate to this feature. HFC is a measure taken across a signal spectrum and can be used to characterize the amount of high-frequency content (HFC) in the signal [12]. As Brossier [13] stated, High Frequency Content (HFC) measure is effective with percussive onsets but less successful determining non-percussive or legato phrases. Then, HFC is sensitive for abrupt onsets but not enough sensitive to the changes of fundamental frequency caused by the left hand.

Aubioonset library [14] gave us the opportunity to tune the peak-picking and silence threshold. One of the key stages of candidate extraction is to optimize peak-picking and silence thresholds in a way that only the plucking hand onsets are determined and the pitch changes due to legato

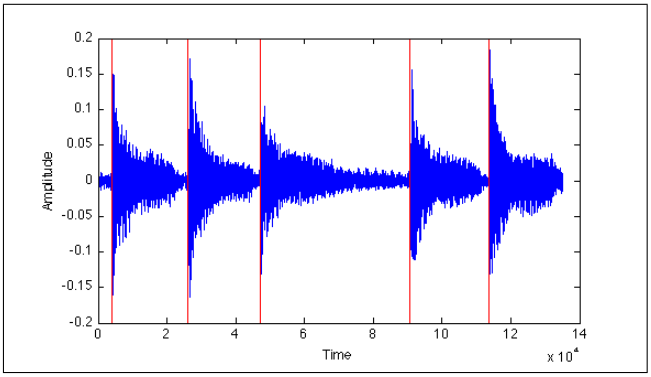


Figure 2: High Frequency Content onsets from the Region Extraction module.

or glissando are not determined as onsets. In order to find suitable parameters for this goal, before running our model, we used a set of hand annotated recordings. Our set is a concatenated audio file which contains 24 non-expressive notes, 6 glissando and 6 legato notes. We hand annotated the onsets of non-expressive, legato and glissando notes. What we want to obtain from Aubioonset was the onsest of the plucking hand. Since this annotated set contains the exact places of onset that we want to obtain from Aubioonset, this set can be considered as our ground truth. After testing with different parameters, we achieved the best results with the following values for algorithm parameters: 1 for peak-picking threshold and $-85db$ for silence threshold.

An example of the resulting onsets proposed by HFC is shown in Figure 2. Specifically, in the exemplified recording six notes are played following the pattern detailed in experiments (see Figure 16) where only 5 of them are plucking onsets. In Figure 2 detected onsets are marked as vertical lines. Between third and fourth detected onsets an expressive articulation (legato) is present. Thus, HFC succeeds because it only determines the onsets caused by the right hand.

The second task performed by the extraction module is to analyze the sound fragment between two onsets. First, each portion between two plucking onsets is analyzed individually. Specifically, two points are determined: the *end*

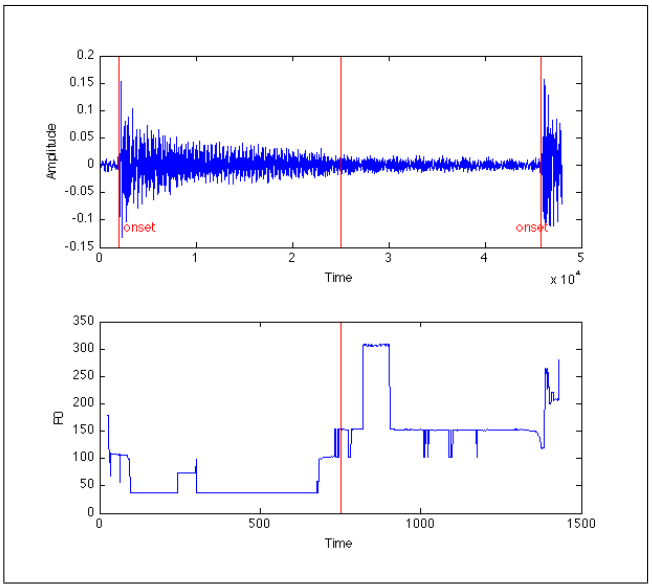


Figure 3: Example of detection of a candidate to an expressive articulation.

of the attack and the *release start*. From experimental measures, we determined attack finish position as $10ms$ after the amplitude reaches its local maximum. We determined the release start position as the final point where local amplitude is equal or greater than 3 percent of the local maximum. Only the fragment between these two points is considered for the further analysis because the noisiest part of a signal is the attack part and the release part of a signal contains unnecessary information for pitch detection (see [15] for details).

We use additional algorithms with a lower threshold in order to capture the changes in fundamental frequency inside the sound fragment. Specifically, complex domain algorithm [16] is used to determine the peaks and Yin [17] is used for the fundamental frequency estimation. Figure 3 shows fundamental frequency evolution between the central region presented in Figure 2. The change of frequency detected points out a possible candidate of expressive articulation. More details can be found in [10].

2.2 Selecting a Descriptor

After extracting the regions candidates to contain expressive articulations, the next step was to analyze them. Because different expressive articulations (legato vs glissando) should present different characteristics in terms of changes in amplitude, aperiodicity, or pitch[8], we focused the analysis on comparing these deviations.

We built representations of these three features (amplitude, aperiodicity, and pitch). Representations helped us to compare different data with different length and density. As we stated above, we are mostly interested in changes: changes in High Frequency Content, changes in fundamental frequency, changes in amplitude, etc. Therefore, we explored the peaks in the examined data because peaks are the points where changes occur.

As an example, Figures 4 and 5 show, from top to bottom, amplitude evolution, pitch evolution, and changes in aperiodicity. As both Figures show, *glissando* and *legato* examples, the changes in pitch are similar. However, the changes in amplitude and aperiodicity present a characteristic slope.

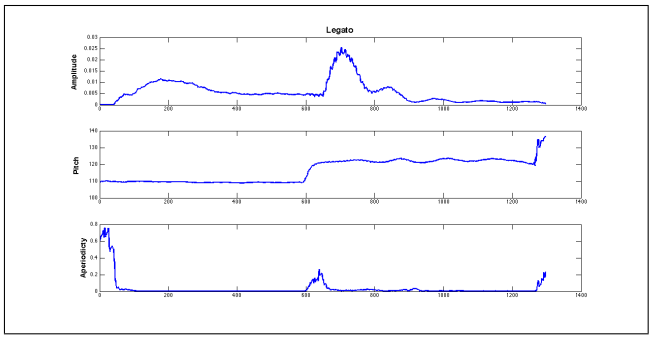


Figure 4: Different features of a legato example. From top the bottom, representations of amplitude, pitch and aperiodicty of the examined legato region.

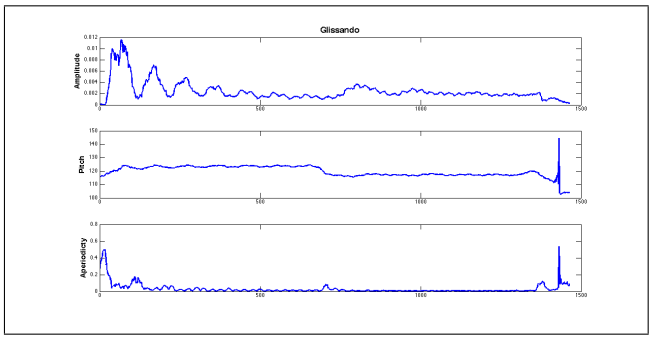


Figure 5: Different features of a glissando example. From top the bottom, representations of amplitude, pitch and aperiodicty of the examined Glissando region.

So, as a first step we concentrated on determining which descriptor could be used. To make this decision, we built models for both aperiodicty and amplitude by using a set of training data. The details of this model construction will be explained in Section 2.4. As a result, we obtained two models (for amplitude and aperiodicity) for both legato and glissando as is shown in Figure 6 and Figure 7. Analyzing the results, amplitude is not a good candidate because the models behave similarly. In contrast, aperiodicity models present a different behavior. Therefore, we selected aperiodicity as the descriptor.

2.3 Preprocessing

Before analyzing and testing our recordings, we applied two different preprocessing techniques to the data in order to make them smoother and ready for comparison.

2.3.1 Smoothing

As expected, aperiodicity portion of the audio file that we are examining includes noise. Our first concern was to avoid this noise and obtain a nicer representation. In order to do that first we applied a 50 step running median smoothing. Running median smoothing is also known as

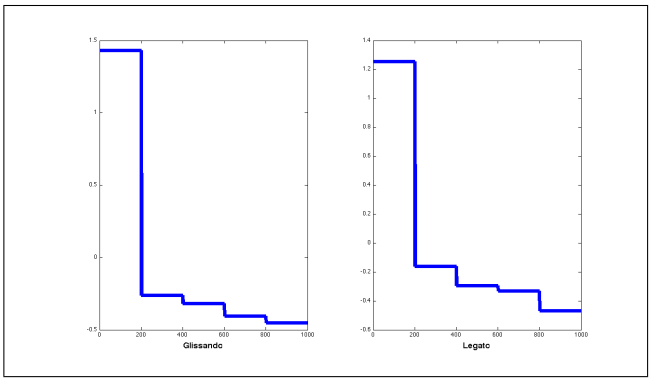


Figure 6: Amplitude models of glissando and legato.

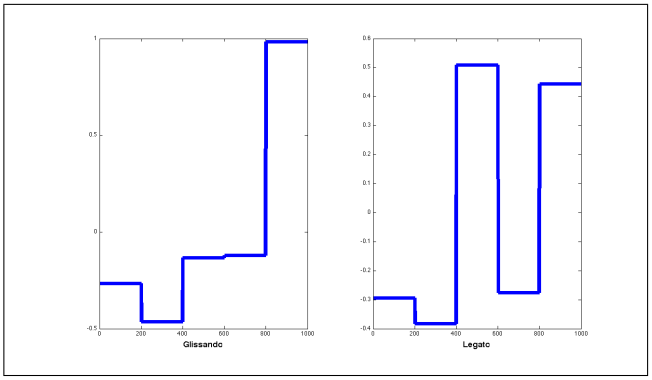


Figure 7: Aperiodicity models of glissando and legato.

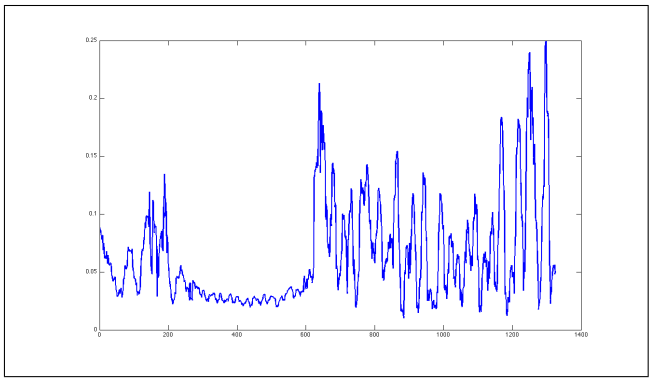


Figure 8: Aperiodicity .

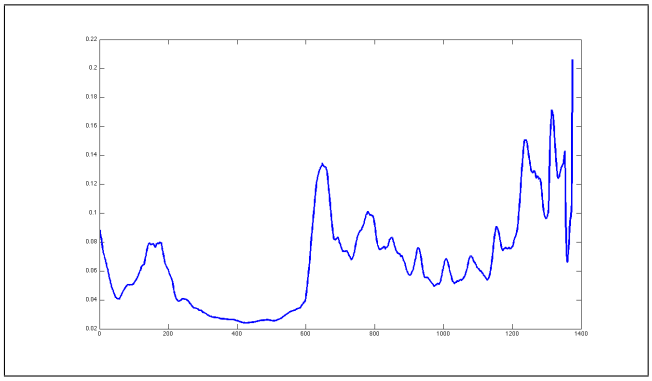


Figure 9: Smoothed Aperiodicity.

median filtering. Median filtering is widely used in digital image processing because under certain conditions, it preserves edges whilst removing noise. In our situation since we are interested in the edges and in removing noise, this approach fits our purposes. By smoothing, the peaks locations of the aperiodicity curves become more easy to extract. Figure 8 and Figure 9 exemplify the smoothing process and show the results we pursued.

2.3.2 Envelope Approximation

After obtaining a smoother data, an envelope approximation algorithm was applied. The core idea of the envelope approximation is to obtain a fixed length representation of the data, specially considering the peaks and also avoiding small deviations by connecting these peak approximations linearly. The envelope approximation algorithm has three parts: *peak peaking*, *scaling* of peak positions according to a fixed length, and *linearly connecting* the peaks. After the envelope approximation, all the data regions we are investigating had the same length, i.e. regions were compressed or enlarged depending their initial size.

We collect all the peaks above a pre-determined threshold. Next, we scale all these peak positions. For instance, imagine that our data includes 10000 bins and we want to scale this data to 1000. And lets say, our peak positions are : 1460, 1465, 1470, 1500 and 1501. What our algorithm does is to scale these peak locations dividing all peak locations by 10 (since we want to scale 10000 to 1000) and round them. So they become 146, 146, 147, 150 and 150. As seen, we have 2 peaks in 146 and 150. In order to fix this duplicity, we choose the ones with the highest peak. After collecting and scaling peak positions, the peaks are linearly connected. As shown in Figure 10, the obtained graph is an approximation of the graph shown in Figure 9. Linear approximation helps the system to avoid consecutive small tips and dips.

In our case all the recordings were performed at 60bpm and all the notes in the recordings are 8th notes. That is, each note is half a second, and each legato or glissando portion is 1 second. We recorded with a sampling rate of 44100, and we did our analysis by using a hop size of 32 bins, i.e. 44100/32 = 1378 bins. We knew that this was our highest limit. For the sake of simplicity, we scaled our x-axis to 1000 bins.

2.4 Building the Models

After applying the preprocessing techniques, we obtained equal length aperiodicity representations of all our expressive articulation portions. Next step was to construct models for both legato and glissando by using this data. In this section we describe how we constructed the models cited briefly in the Section 2.2 (and shown in Figure 6 and Figure 7). The following steps were used to construct the models: *Histogram Calculation*, *Smoothing* and *Envelope approximation* (explained in Section 2.3), and finally, SAX representation. In this section we present the Histogram Calculation and the SAX representation techniques.

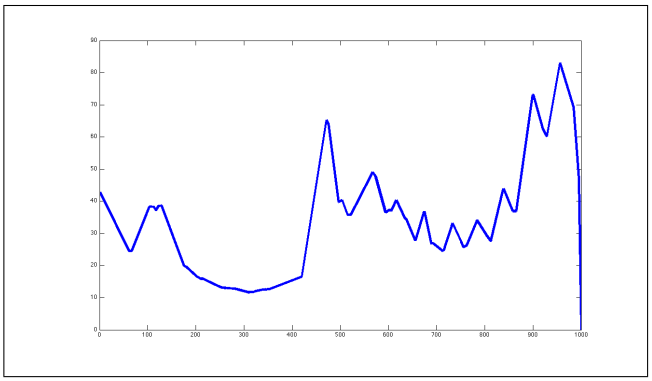
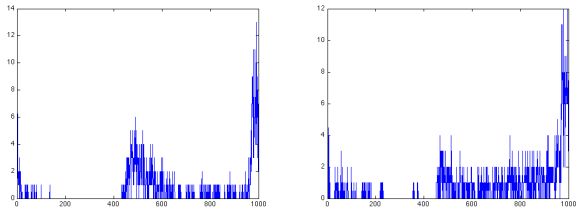


Figure 10: Envelope approximation of a legato portion.



(a) Legato Histogram (b) Glisando Histogram

Figure 11: Peak histograms of our legato and glissando training sets.

2.4.1 Histogram Calculation

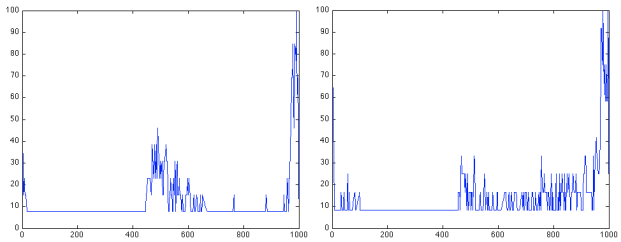
Another method that we are using is histogram envelope calculation. We use this technique to calculate the peak density of a set of data. Specifically, a set of recordings containing 36 legato and 36 glissando examples (recorded by a professional classical guitarist) was used as training set. First, for each legato and glissando example, we determined the peaks. Since we want to model the places where condensed peaks occur, this time we use a threshold which is 30 percent and collect the peaks which have amplitude values above this threshold. Notice that the threshold is different than the used in envelope approximation. Then, we used histograms to compute the density of the peak locations. Figure 11 shows the resulting histograms.

After constructing the histograms, as shown in Figure 11, we used our envelope approximation method to construct the envelopes of legato and glissando histogram models (see Figure 12).

2.4.2 SAX: Symbolic Aggregate Approximation

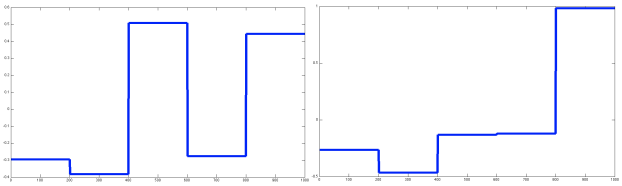
Although the histogram envelope approximations of legato and glissando in Figure 12 are close to our purposes, they still include noisy sections. Rather than these abrupt changes (noises), we are interested in a more general representation reflecting the changes more smoothly.

SAX (Symbolic Aggregate Approximation) [18], is a symbolic representation used in time series analysis that provides a dimensionality reduction while preserving the properties of the curves. Moreover, SAX representation makes the distance measurements easier. Then, we applied



(a) Legato Final Envelope (b) Glisando Final Envelope

Figure 12: Final envelope approximation of peak histograms of legato and glissando training sets.



(a) Legato SAX Representation (b) Glisando SAX Representation

Figure 13: SAX representation of legato and glissando final models.

the SAX representation to histogram envelope approximations.

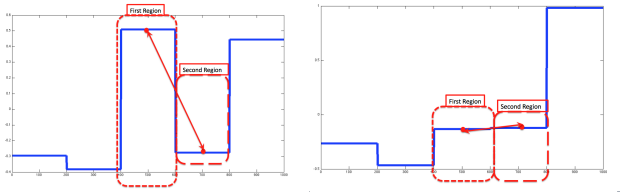
As we mentioned in Section 2.3.2, we scaled the x-axis to 1000. We made tests with step sizes of 10 and 5. As we report in the Experiments section, an step size of 5 gave better results. We also tested with step sizes lower than 5, but the performance clearly decreased. Since we are using an step size of 5, each step becomes 100 bins in length. After obtaining the SAX representation of each expressive articulation, we used our distance calculation algorithm which we are going to explain in the next section.

2.5 Detection

After obtaining the Sax representation of our glissando and legato models, we divided them into 2 regions, a first region between bins 400 and 600, and a second region between bins 600 and 800 (see Figure 14).

For the expressive articulation excerpt, we have the envelope approximation representation with the same length of the SAX representation of final models. So, we can compare the regions. For the final expressive articulation models (see Figure 13) we took the value for each region and compute the deviation (slope) between these two regions. We make this computation for both legato and glissando models separately.

We also compute the same deviation for each expressive articulation envelope approximation (see Figure 15). But this time, since we do not have SAX representation, for each region we do not have single values. Therefore, for each region we compute the local maxima and take the deviation (slope) of these two local maxima. After obtaining this value, we compare it with the numbers that we obtained from both final models of legato and glissando. If



(a) Legato (b) Glisando

Figure 14: Peak occurrence deviation.

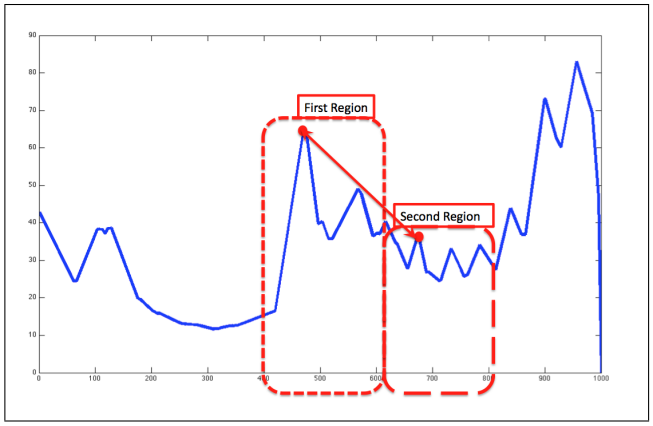


Figure 15: Expressive articultion difference.

the deviation value is closer to the legato model, we annotate this expressive articulation as a legato and vice versa.

3. EXPERIMENTS

The goal of the experiments was to test the accuracy of our approach. Because legato and glissando can be played in ascending or descending intervals, we were also interested in studying the results considering these two movements. Additionally, since in a guitar there are three nylon and three metallic strings, we also studied the results on these two sets of strings.

Borrowing from Carlevaro's guitar exercises [19], we recorded a collection of ascending and descending chromatic scales. Legato and glissando examples were recorded by a professional classical guitar performer. The performer was asked to play chromatic scales in three different regions of the guitar fretboard. Specifically, we recorded notes from the first 12 frets of the fretboard where each recording concentrated in 4 specific frets. The basic exercise from the first fretboard region is shown in Figure 16. Each scale contains 24 ascending and 24 descending notes. Each exercise contains 12 expressive articulations (the ones connected with an arch). Since we repeated the exercise at three different positions, we obtained 36 legato and 36 glissando examples.

We presented all the 72 examples to our system. Then, our system proposed a possible expressive articulation as described in Section 2. Results are reported in Table 1.

First, we may observe that a step size of 5 is the most appropriate setting. This result corroborates that a higher resolution when discretizing is not required and demonstrates

	Step Size	
Recordings	**5**	**10**
Ascending Legato	100%	100%
Descending Legato	66.6%	72.2%
Ascending Glissando	83.3%	61.1%
Descending Glissando	77.7%	77.7%
Glissando Metallic Strings	77.7%	77.7%
Glissando Nylon Strings	83.3%	61.1%
Legato Metallic Strings	86.6%	80%
Legato Nylon Strings	73.3%	86.6%

Table 1: Performance of our model applied to test set

that the SAX representation provides a powerful technique to summarize the information about changes.

The overall performance for legato identification is 83.5%. Notice that ascending legato reaches a 100% of accuracy whereas descending legato achieves a 66.6%. Regarding glissando, there is no difference between ascending or descending accuracy (83.3%,77.7%). Finally, analyzing the results when considering the string type, the results show a similar accuracy.

4. CONCLUSION

In this paper we presented some preliminary results on the identification of two classical guitar articulations, legato and glissando, from a collection of chromatic exercises recorded by a professional guitarist. Our approach uses aperiodicity information to identify the articulation and a SAX representation to characterize the articulation models.

Reported results show that our system is able to differentiate successfully among these two articulations. Our next goal is to study the capabilities of our approach in the context of a real performance. To avoid the analysis on a polyphonic recording, we plan to use an hexaphonic pickup.

5. ACKNOWLEDGMENTS

This work was partially funded by NEXT-CBR (TIN2009-13692-C03-01), IL4LTS (CSIC-200450E557) and by the Generalitat de Catalunya under the grant 2009-SGR-1434. Tan Hakan Özaslan is a Phd student of the Doctoral Program in Information, Communication, and Audiovisuals Technologies of the Universitat Pompeu Fabra.

6. REFERENCES

[1] P. Juslin, "Communicating emotion in music performance: a review and a theoretical framework," in *Music and emotion: theory and research* (P. Juslin and J. Sloboda, eds.), pp. 309–337, New York: Oxford University Press, 2001.

[2] E. Lindström, "5 x "oh, my darling clementine". the influence of expressive intention on music performance," 1992. Department of Psychology, Uppsala University.

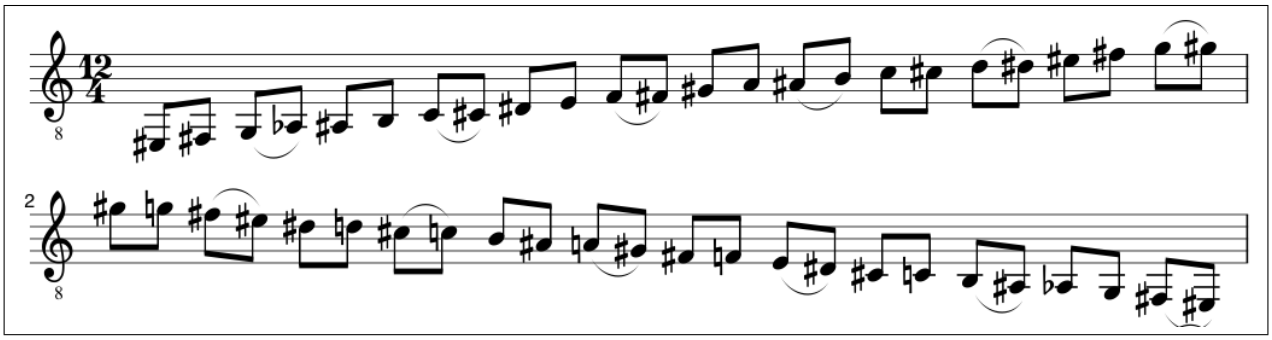


Figure 16: Legato Score in first position.

[3] A. Gabrielsson, "Expressive intention and performance," in *Music and the Mind Machine* (R. Steinberg, ed.), pp. 35–47, Berlin: Springer-Verlag, 1995.

[4] J. A. Sloboda, "The communication of musical metre in piano performance," *Quarterly Journal of Experimental Psychology*, vol. 35A, pp. 377–396, 1983.

[5] A. Gabrielsson, "Once again: The theme from Mozart's piano sonata in A major (K. 331). A comparison of five performances," in *Action and perception in rhythm and music* (A. Gabrielsson, ed.), pp. 81–103, Stockholm: Royal Swedish Academy of Music, 1987.

[6] C. Palmer, "Anatomy of a performance: Sources of musical expression," *Music Perception*, vol. 13, no. 3, pp. 433–453, 1996.

[7] C. Erkut, V. Valimaki, M. Karjalainen, and M. Laurson, "Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar," in *108th AES Convention*, pp. 19–22, February 2000.

[8] J. Norton, *Motion capture to build a foundation for a computer-controlled instrument by study of classical guitar performance*. PhD thesis, Stanford University, September 2008.

[9] H. Heijink and R. G. J. Meulenbroek, "On the complexity of classical guitar playing:functional adaptations to task constraints," *Journal of Motor Behavior*, vol. 34, no. 4, pp. 339–351, 2002.

[10] T. Ozaslan, E. Guaus, E. Palacios, and J. Arcos, "Attack based articulation analysis of nylon string guitar," in *CMMR 2010*, June 2010.

[11] C. Traube and P. Depalle, "Extraction of the excitation point location on a string using weighted least-square estimation of a comb filter delay," in *In Procs. of the 6th International Conference on Digital Audio Effects (DAFx-03*, 2003.

[12] P. Masri, *Computer modeling of Sound for Transformation and Synthesis of Musical Signal*. PhD thesis, University of Bristol, 1996.

[13] P. Brossier, J. P. Bello, and M. D. Plumbley, "Real-time temporal segmentation of note objects in music signals," in *Proceedings of the International Computer Music Conference (ICMC2004)*, November 2004.

[14] P. Brossier, *Automatic annotation of musical audio for interactive systems*. PhD thesis, Centre for Digital music, Queen Mary University of London, 2006.

[15] C. Dodge and T. A. Jerse, *Computer Music: Synthesis, Composition, and Performance*. Macmillan Library Reference, 1985.

[16] C. Duxbury, J. Bello, J. Davies, S. M., and M. Mark, "Complex domain onset detection for musical signals," in *Proceedings Digital Audio Effects Workshop*, 2003.

[17] A. de Cheveigné and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.

[18] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing sax: a novel symbolic representation of time series," *Data Mining and Knowledge Discovery*, vol. 15, pp. 107–144, October 2007.

[19] A. Carlevaro, "Serie didactica para guitarra," vol. 4, Barry Editorial, 1974.