Attack Based Articulation Analysis of Nylon String Guitar

Tan Hakan Özaslan, Enric Guaus, Eric Palacios, and Josep Lluis Arcos

IIIA, Artificial Intelligence Research Institute
CSIC, Spanish National Research Council
Campus UAB, 08193 Bellaterra, Spain
{tan,eguaus,epalacios,arcos}@iiia.csic.es

Abstract. The study of musical expressivity is an active field in sound and music computing. The research interest comes from different motivations: to understand or model music expressivity; to identify the expressive resources that characterize an instrument, musical genre, or performer; or to build synthesis systems able to play expressively. In this paper, we present a system that focuses on the study of expressivity in nylon-string guitars. Specifically, our system combines several state of the art analysis algorithms to identify guitar left-hand articulations such as legatos and appoggiaturas. We describe the components of our system and provide some preliminary results by analyzing single articulations and some short melodies.

1 Introduction

Expressivity can be described as the differences (deviations) between a musical score and its execution. These deviations are mainly motivated by two purposes: to clarify the musical structure [1–3] and as a way to communicate affective content [4–6]. Moreover, these expressive deviations vary depending on the musical genre, the instrument, and the performer. Specifically, each performer has his/her own unique way to add expressivity by using the instrument.

Guitar is the one of the most popular instruments in western music. Thus, most of the music genres include the guitar. Although plucked instruments and guitar synthesis have been studied extensively [7–9], expressive articulation analysis from real guitar recordings has not been fully tackled. This analysis is complex because guitar is an instrument with a rich repertoire of expressive articulations and because, either when playing guitar melodies, several strings may be vibrating at the same time. As an additional statement, even synthesis of a single tone is a complex subject [7].

The first step when analyzing guitar expressivity is to identify and characterize the way notes are played, i.e. guitar articulations. The analysis of expressive articulations has been previously performed with image analysis techniques. In his dissertation Norton [9] proposed the use of a motion caption system based on PhaseSpace Inc. Moreover, Burns and Heijink proposed different methods for analyzing left-hand fingerings of a classical guitar. Heijink and Meulenbroek [10] used a three-dimensional motion tracking system. Burns and Wanderley [11] proposed a method to visually detect and recognize fingering gestures.

In guitar playing both hands are used: one hand is used to press the strings on the fretboard 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, fretting hand is also involved in the creation of a note onset or different expressive articulations such as legatos, appoggiaturas, glissandi, and vibratos.

Some guitarists use the right hand to pluck the strings whereas others use the left hand. For sake of simplicity, in the rest of the document we consider the hand that plucks the strings as the right hand and the hand that presses the frets as the left hand.

According to Norton [9], guitar articulations can be divided into three main groups related to the place of the sound where they act: *attack*, *sustain*, and *release* articulations. In this research we are focusing on the identification of attack articulations such as legatos and appoggiaturas. The technique used to play legatos differs depending on whether is an ascending or a descending legato. An ascending legato (also known as hammer-on) is achieved by fretting a note with a left-hand finger. A descending legato (also known as pull-off) is achieved by plucking the string with a left-hand finger currently used to play a previous note. Legatos are notated with a slur symbol. An appoggiatura (notated with a grace note) is an expressive articulation where a short note is added, one degree higher or lower than the principal note, before the principal note. In guitar this expressive resource is achieved by sliding a left-hand finger from one note to another.

In this paper we present an automatic detection system from audio recordings. Our system is mainly based on a combination of different onset detection algorithms. The structure of the paper is as follows: Section 2 describes our methodology for articulation determination. Section 3 focuses on the experiments conducted to evaluate our approach. Last section, Section 4, summarizes current results and presents the next research steps.

2 Methodology

Articulation refers to how the pieces of something are joined together. In music, these pieces are the notes and the different ways of manipulating them are called articulations. In this paper we propose a system able to identify left-hand articulations such as legatos and appoggiaturas. In order to achieve our goal, we combined the information obtained from several audio analysis algorithms.

Our approach is based on first determining the note onsets caused when plucking the strings. Next, a more finely grained analysis is performed inside the regions delimited by two plucking onsets. A simple representation diagram of our model is shown in Figure 1.

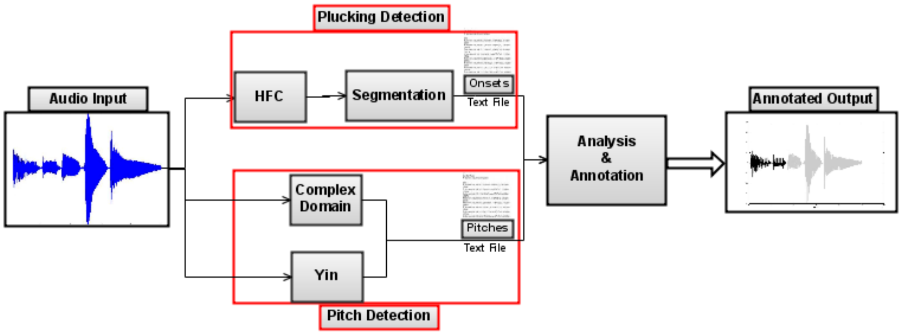


Fig. 1: Main diagram of our model.

For this analysis we used Aubio library [12]. Aubio is a library collecting a collection of state of the art algorithms aimed at annotating audio signals. Aubio library includes four main applications: *aubioonset*, *aubionotes*, *aubiocut*, and *aubiopitch*. Each application gives us the chance of trying different algorithms and also of tuning several other parameters. In the current prototype we are using *aubioonset* for our plucking detection module and *aubionotes* for our pitch detection module.

At the end we combine the outputs from both modules and decide whether there is an expressive articulation or not. In the next two sections the plucking detection module and the pitch detection module are described. Finally, in Section 2.3 we explain how we combine the information provided by these two modules to determine the existence of expressive articulations.

2.1 Plucking Detection

Our first task is to determine the onsets caused by the plucking hand. As we stated before, guitar performers can apply different articulations by using both of their hands. However, the kind of articulations that we are investigating (legatos and appoggiaturas) are performed by the left hand. Although they can cause onsets, these onsets are not as powerful in terms of both energy and harmonicity [13]. Therefore, we need an onset determination algorithm suitable to this specific characterictic.

The High Frequency Content measure (HFC) [14] is a measure taken across a signal spectrum that can be used to characterize the amount of high-frequency content in the signal. As Brossier stated, HFC is effective with percussive onsets but less successful determining non-percussive and legato phrases [15]. As right-hand onsets are more percussive than left-hand onsets, HFC was the strongest candidate for the plucking detection algorithm, because is sensitive to abrupt onsets but not enough sensitive to the changes of fundamental frequency caused by left-hand articulations. Thus, this is the main reason we chose HFC to determine plucking onsets.

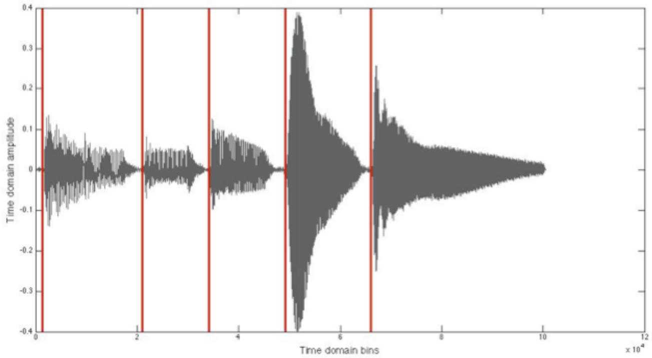


Fig. 2: HFC onsets.

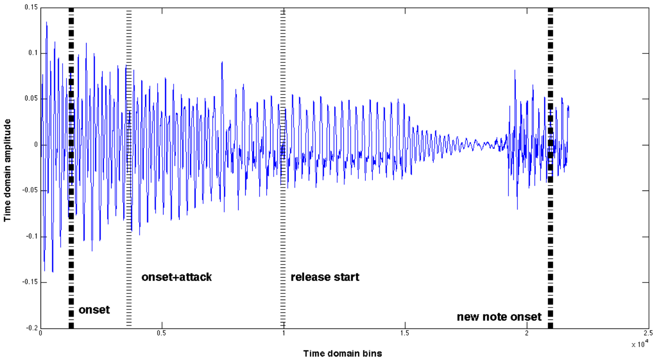


Fig. 3: Features of the portion between two onsets.

Aubioonset library gave us the opportunity to tune the peak-picking threshold, which we tested with a set of hand labeled recordings, including both articulated and non-articulated notes. We used 1.7 for peak picking threshold and $-95db$ for silence threshold. We used this set as our ground truth and tuned our values according to this set.

An example of the resulting onsets proposed by HFC is shown in Figure 2. Specifically, in the exemplified recording 5 plucking onsets are detected (noted with vertical lines). HFC succeeds because only onsets caused by the plucking hand are detected. Moreover, between some of two detected onsets expressive articulations are present. As expected, these articulations are not detected as plucking onsets.

Next, each portion between two plucking onsets is analyzed individually. Specifically, we are interested in determining two points: the *end of the attack* and the *release start*. Because, the portion between these points contains the most valuable information for pitch detection [16]. From experimental measures, we determine attack finish position $10ms$ after the amplitude reaches its local maximum.

Determining the release point was more difficult. In guitar sustain portions, which include the most valuable information for pitch detection, and release portions are not separated so distinctively like a key instrument. In key instruments like a piano, the sustain portion is the time where player keeps pressing the keys, and release portion starts when the player hold-offs the keys. However in guitar there is only one pluck for a note and the places where sustain ends and release starts are not obvious. In order to determine the release start position we have used a percent measure rather than an absolute threshold. Specifically, we determine the release starting point as the point where local amplitude is equal or greater than 3 percent of the local maximum.

For example, in Figure 3, the first portion of the Figure 2 is zoomed. The first and the last lines are the plucking onsets identified by HFC algorithm. The first dashed line is the place where attack finishes. The second dashed line is the place where release starts.

2.2 Pitch Detection

Our second task was to analyze the sound fragment between two onsets. Since we know the onset values of plucking hand, a peak detection algorithm with a lower threshold is required in order to capture the changes in fundamental frequency. Specifically, if fundamental frequency is not constant between two onsets, we consider that the possibility of the existence of an expressive articulation is high.

In the pitch detection module, i.e to extract onsets and their corresponding fundamental frequencies, we used *aubionotes*. In Aubio library, both onset detection and fundamental frequency estimation algorithms can be chosen from a bunch of alternatives. At this stage, we used complex domain algorithm [17] to determine the peaks and Yin [18] for the fundamental frequency estimation because we require a more sensitive algorithm than the one used to detect the plucking onsets.

Complex domain onset detection is based on a combination of phase approach and energy based approach. The algorithm parameters were fixed to 2048 bins as window size, 512 bins as hop size, 1 as pick peaking threshold, and $-95db$ as silence threshold. Using these parameters we obtained an output like the one shown in Figure 4. As shown in the figure, first results were not as we expected. Specifically, they were noisier than expected. There were noisy parts, especially at the beginning of the notes, which generated false-positive peaks. For instance in Figure 4, many false positive note onsets are detected between the interval from 0 to 0.2 seconds.

A careful analysis of the results demonstrated that the false-positive peaks were located in the region of the notes frequency borders. Therefore, we propose a lightweight solution for the problem: to apply a chroma filtering to the regions

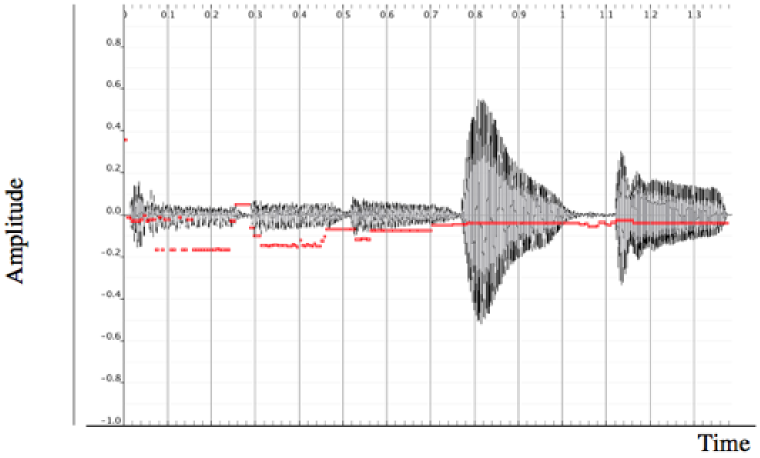


Fig. 4: Note Extraction without chroma feature.

that are in the borders of complex domain peaks. As shown in Figure 5, after applying chroma conversion, the results are drastically improved.

Next, we analyzed the fragments between two onsets based on the segments provided by the plucking detection module. Specifically, we analyzed the sound fragment between attack ending point and release starting point (because the noisiest part of a signal is the attack part and the release part of a signal contains unnecessary information for pitch detection [16]). Therefore, in this analysis only the fragment between attack and release parts, where pitch information is relatively constant, is used.

Figure 6 shows fundamental frequency values and plucking onsets. X-axis represents the time domain bins and Y-axis represents the frequency. In Figure 6, vertical lines depict the attack and release parts respectively. In the middle there is a change in frequency, which was not determined as an onset by the first module. Although it seems like an error, it is a success result for our model. Specifically, in this phrase there is an appoggiatura, a left-hand articulation, and was not identified as an onset by plucking detection module (HFC algorithm), but identified by the pitch detection module (Complex Domain algorithm). The output of the pitch detection module for this recording is shown in Table 1.

2.3 Analysis and Annotation

After obtaining the results from plucking detection and pitch detection modules, the goal of the analysis and annotation module is to determine the candidates of expressive articulations. Specifically, from the results of the pitch detection module, we analyze the differences of fundamental frequencies in the segments between attack and release parts (provided by the plucking detection module). For instance, in Table 1 the light gray values represent the attack and release parts, which we did not take into account while applying our decision algorithm.

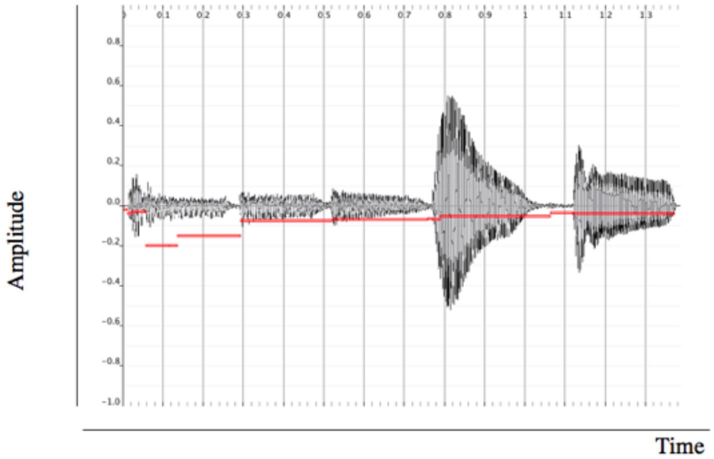


Fig. 5: Note Extraction with chroma feature.

Note Start (ms.)	Fundamental Frequency
0.02	130
0.19	130
0.37	130
0.46	146
0.66	146
0.76	146
099	146
1.10	146
1.41	174
1.48	116

Table 1: Output of the pitch detection module.

The differences of fundamental frequencies are calculated by subtracting to each bin its preceding bin. Thus, when the fragment examined is a non-articulated fragment, this operation returns 0 for all bins. On the other side, in fragments containing expressive articulations some peaks arise (see Figure 7 for an example).

In Figure 7 there is only one peak, but in other recordings some consecutive peaks may arise. The explanation is that the left-hand finger also causes an onset, i.e. it generates a transient part. As a result of this transient, more than one change in fundamental frequency may be present. If those changes or peaks are close to each other we consider them as a single peak. We define this closeness with a pre-determined consecutiveness threshold. Specifically, if the maximum distance between these peaks is 5 bins we consider there is a single peak, i.e. the fragment is a candidate to contain an expressive articulation. However, if

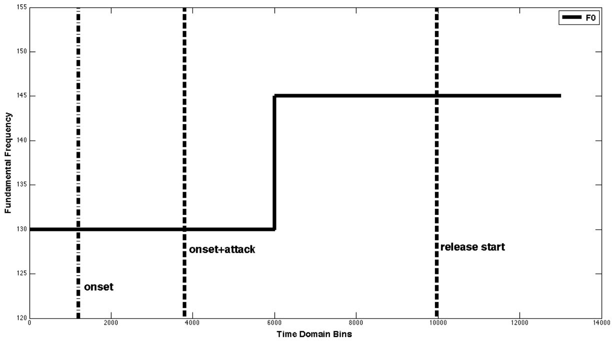


Fig. 6: Example of an appoggiatura articulation.

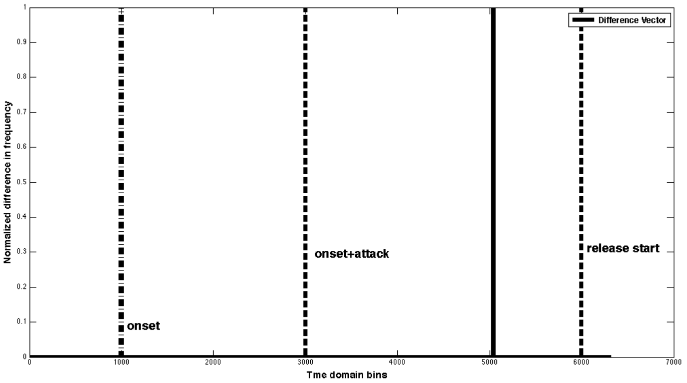


Fig. 7: Difference vector of pitch frequency values of fundamental frequency array.

the peaks are separated each other more than the consecutiveness threshold, the fragment is not considered as an articulation candidate. Our consideration is that these peaks respond to a probable noisy part of the signal, a crackle in the recording, or a digital conversion error.

The final step in our system is to determine the expressive articulation. Because in the current system we are only analyzing appoggiaturas and legatos, measuring the duration of the first sub-segment both articulations are easily differentiated.

We colored the parts where expressive articulations are identified and annotate the articulation. Figure 8 shows the annotation of *Phrase_2*. As summarized in Table 3, *phrase_2* has two expressive articulations. To show the places of these expressive articulations, in the wave representation of the sound, in Figure 8, we colored them black or bold.

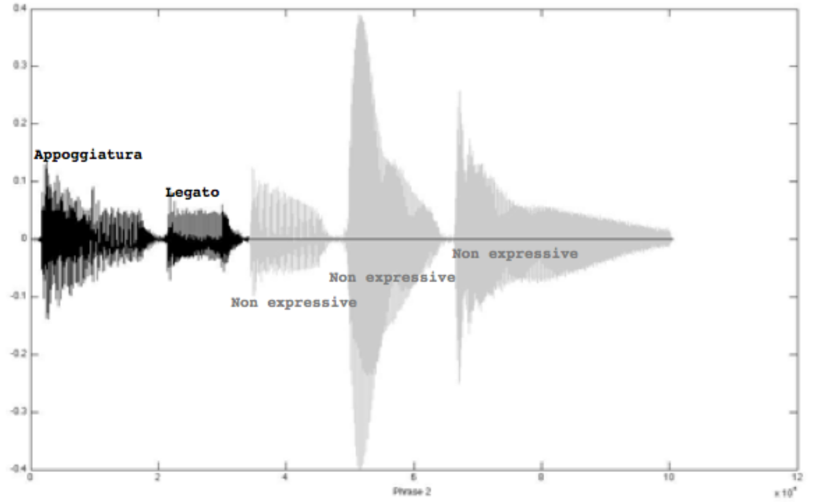


Fig. 8: Annotated output of Phrase_2.

3 Experiments

The goal of the experiments is to test the performance of our model. First, we analyzed the accuracy of our approach in detecting legatos. The hypothesis was that legatos are the articulations easiest to detect because are composed by two long notes. Next, we analyzed the accuracy on appoggiaturas. Because in this situation the first note (the grace note) has a short duration, it can be confused with the attack.

Additionally, in all of the experiments we were interested in studying two different issues: (1) the accuracy of ascending or descending articulations and (2) the accuracy on non-metallic (1st to 3rd guitar strings) and metallic wounded strings (4th to 6th strings). For the first issue, the hypothesis is that the melodic direction should not affect the accuracy. Nevertheless, we expect a lower accuracy in metallic strings because they produce more percussive notes.

We recorded examples using the third (non-metallic) and fourth (metallic) string. For each combination of string-articulation-direction at least 10 examples were used. Notice that we also performed recordings with a neutral articulation (neither legato nor appoggiatura). Totally we had a set of 105 examples of two notes guitar recordings. Moreover, we studied the accuracy of our system in the context of short melodies (including 5-6 notes) where different combinations of expressive articulations were played. Recorded examples and a detailed explanation of the testing set is available at `http://www.iiia.csic.es/guitarLab`.

Recordings	**3rd String**	**4th String**	**Correct/All**
Non-expressive	91.6 %	84.9 %	22/25
Ascending Legatos	80.0 %	90.0 %	17/20
Descending Legatos	90.0 %	70.0 %	16/20
Ascending Appoggiaturas	70.0 %	70.0 %	14/20
Descending Appoggiaturas	80.0 %	70.0 %	15/20

Table 2: Performance of our model applied to single articulations.

Excerpt Name	**Ground Truth**	**Detected**
Phrase_1	1	2
Phrase_2	2	2
Phrase_3	0	0
Phrase_4	2	3
Phrase_5	1	1

Table 3: Results of our model applied to short phrases.

3.1 Basic scenarios

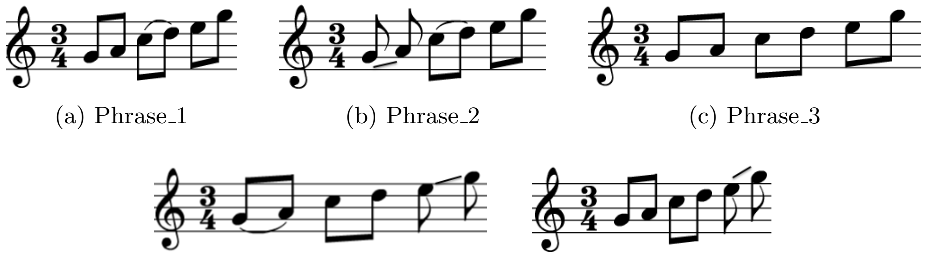
We first applied our system to single expressive and non-expressive articulations. All the recordings were hand labeled; they were also our ground truth. We compared the output results with expected annotations.

Analyzing the experiments (see Table 2), different conclusions can be extracted. First, as expected, legatos are easier to detect than appoggiaturas. Second, in a non-metallic string the melodic direction do not determine a different performance. Regarding a metallic string, descending legatos are more difficult to detect than ascending legatos. This result is not surprising because the plucking action of left-hand fingers in descending legatos is slightly similar to a right-hand plucking. However, this difference does not appear in appoggiaturas because the finger movement is the same.

3.2 Short melodies

As a preliminary test with more realistic recordings, we also recorded a small set of 5-6 note phrases. They include different articulations in random places (see Figure 9). As shown in Table 3, each phrase includes different number of expressive articulations varying from 0-2.

We applied our model to these recordings with the same settings we used with short phrases except for the release threshold. Specifically, since in short phrase recordings the transition parts between two notes have more noise, it increases the average value of the amplitude between two onsets. Because of that, the release threshold in more realistic recordings has to be increased. Specifically, after some preliminary tests, we fixed the release threshold to 30%.



(a) Phrase_1 (b) Phrase_2 (c) Phrase_3

(d) Phrase_4 (e) Phrase_5

Fig. 9: Short melodies.

Analyzing the results, the performance of our model was similar to the previous experiments, i.e. when we analyze single articulations. However, in two phrases where a note was played with a soft right-hand plucking, these notes were proposed as legato candidates. In Figures 9a - 9e all short melodies can be seen. For instance the melody in Figure 8 corresponds to Figure 9b. Phrase_3 which is Figure 9c where there is no expressive articulation and Phrase_4 which is Figure 9d, is the same notes with Phrase_3 but it includes two expressive articulations, first one is a legato and second one is an appoggiatura.

4 Conclusions

In this paper we presented a system to identify left-hand articulations such as legatos and appoggiaturas. Our approach combined the audio information extracted using several existing audio analysis algorithms. Specifically, we have used HFC for plucking detection and Complex Domain and YIN algorithms for pitch detection. Then, combining the data coming from these two sources, we developed a first decision mechanism to identify attack articulations.

Although we are aware that our current system may be improved, the results show that it is able to identify successfully these two attack-based articulations in non-metallic strings. As expected legatos are easier to identify than appoggiaturas. Specifically, the short duration of appoggiaturas is sometimes confused as a single note attack.

We are currently working in improving the performance for metallic strings. Specifically, we are exploring the possibility of dynamically changing the parameters of the analysis algorithms. For instance, depending on the string where notes are played, use different analysis thresholds.

As a next step, we plan to incorporate more analysis and decision components into our system with the aim of covering all the main expressive articulations used in guitar playing.

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.

References

1. Sloboda, J.A.: The communication of musical metre in piano performance. Quarterly Journal of Experimental Psychology **35A** (1983) 377–396
2. Gabrielsson, A.: Once again: The theme from Mozart's piano sonata in A major (K. 331). A comparison of five performances. In Gabrielsson, A., ed.: Action and perception in rhythm and music. Royal Swedish Academy of Music, Stockholm (1987) 81–103
3. Palmer, C.: Anatomy of a performance: Sources of musical expression. Music Perception **13**(3) (1996) 433–453
4. Juslin, P.: Communicating emotion in music performance: a review and a theoretical framework. In Juslin, P., Sloboda, J., eds.: Music and emotion: theory and research. Oxford University Press, New York (2001) 309–337
5. Lindström, E.: 5 x "oh, my darling clementine". the influence of expressive intention on music performance (1992) Department of Psychology, Uppsala University.
6. Gabrielsson, A.: Expressive intention and performance. In Steinberg, R., ed.: Music and the Mind Machine. Springer-Verlag, Berlin (1995) 35–47
7. Erkut, C., Valimaki, V., Karjalainen, M., Laurson, M.: Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar. In: 108th AES Convention. (February 2000) 19–22
8. Laurson, M., Erkut, C., Välimäki, V., Kuuskankare, M.: Methods for modeling realistic playing in acoustic guitar synthesis. Comput. Music J. **25**(3) (2001) 38–49
9. Norton, J.: Motion capture to build a foundation for a computer-controlled instrument by study of classical guitar performance. PhD thesis, Stanford University (September 2008)
10. Heijink, H., Meulenbroek, R.G.J.: On the complexity of classical guitar playing:functional adaptations to task constraints. Journal of Motor Behavior **34**(4) (2002) 339–351
11. Burns, A., Wanderley, M.: Visual methods for the retrieval of guitarist fingering. In: NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression. (June 2006) 196–199
12. Brossier, P.: Automatic annotation of musical audio for interactive systems. PhD thesis, Centre for Digital music, Queen Mary University of London (2006)
13. Traube, C., Depalle, P.: 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)
14. Masri, P.: Computer modeling of Sound for Transformation and Synthesis of Musical Signal. PhD thesis, University of Bristol (1996)

15. Brossier, P., Bello, J.P., Plumbley, M.D.: Real-time temporal segmentation of note objects in music signals. In: Proceedings of the International Computer Music Conference (ICMC2004). (November 2004)
16. Dodge, C., Jerse, T.A.: Computer Music: Synthesis, Composition, and Performance. Macmillan Library Reference (1985)
17. Duxbury, C., Bello, J., Davies, J., M., S., Mark, M.: Complex domain onset detection for musical signals. In: Proceedings Digital Audio Effects Workshop. (2003)
18. de Cheveigné, A., Kawahara, H.: Yin, a fundamental frequency estimator for speech and music. The Journal of the Acoustical Society of America **111**(4) (2002) 1917–1930