ANALYZING LEFT HAND FINGERING IN GUITAR PLAYING

Enric Guaus, Josep Llus Arcos
Artificial Intelligence Research Institute, IIIA.
Spanish National Research Council, CSIC.
{eguaus,arcos}@iiia.csic.es

ABSTRACT

In this paper, we present our research on left hand gesture acquisition and analysis in guitar performances. The main goal of our research is the study of expressiveness. Here, we focus on a detection model for the left hand fingering based on gesture information. We use a capacitive sensora to capture fingering positions and we look for a prototypical description of the most common fingering positions in guitar playing. We report the performed experiments and study the obtained results proposing the use of classification techniques to automatically determine the finger positions.

1. INTRODUCTION

Guitar is one of the most popular instruments in western culture. The guitar (and the music it produces) has been object of study in many disciplines, i.e. musicology, sociology, physics or computer science. Focusing on acoustic and signal processing disciplines, there are many interesting studies explaining its physical behavior and produced sound [1, 2]. Nevertheless, the essence of guitar music is sometimes reflected by subtile particularities which are completely dependent on the players, styles, and musical genres. Although some successful approaches exist in the literature [3], these particularities are, sometimes, difficult to identify only with recorded audio data. The richness of the guitar expressivity raises a challenge that, even analyzing each string individually, i.e. using hexaphonic pickups, it is still partially tackled. In this context, caption of gestures in guitar performances becomes a good complement to the audio recording.

The study of performer gestures in music is not new. For instance, Young [4] presented a system to capture the performance parameters in violin playing. Focusing on the guitar, there are some interesting approaches studying the gestures of guitar players [5, 6]. Centering on the finger movements, the available approaches are traditionally based on the analysis of images. Burns and Wanderley [7, 8] proposed a method to visually detect and recognize fingering gestures of the left hand of a guitarist. Heijink and Meulenbroek [9] proposed the use of a three-dimensional motion tracking system (Optotrak 3020) to analyze the behavior of the left hand in a classical guitar. Norton [10] proposed the use of another optical motion caption system based on the Phase Space Inc., with quite successful results. Although these optical systems have proved to partly solve and represent guitar gestures, some occlusion problems may appear in specific finger positions. The proposed acquisition system is a good complement to the existing ones.

Our research focuses on understanding of particular articulations used by different players, styles or musical genres. For that, we need to capture gesture information from the left hand and to detect its exact position. With such information, we can (1) detect the fingering in a given score, and (2) predict the possible articulations and plucked strings even before the sound is produced. The goal of this paper is to present a model that detects the left hand position, based on gesture information, using classification techniques.

The paper is organized as follows: First, in Section 2, we describe the sensors we use. Then, Section 3 shows the list of recorded excerpts, and explains the pattern creation process from the recorded data. Next, in Section 4, we carefully analyze the obtained recordings, propose the use of classification techniques to automatically classify the patterns, and analyze the results. Finally, we summarize the results achieved, present research conclusions, and propose the next steps of our research in Section 5.

2. ACQUISITION

The acquisition system is based on capacitive sensors, described in [11]. Capacitive sensors are not new. In 1919, Lev Termen invented the Theremin, considered the first electronic instrument in history. Lev Termen exploited the capacitive effect of a player near two antennas, one controlling the pitch and the other controlling the loudness, of an harmonic signal. More recently, new musical interfaces also use capacitive sensors to control musical parameters [12, 13].

The proposed system consists of an array of capacitive sensors, mounted on the fretboard of the guitar, configured in *load mode* [14], where the distance between the electrode and a single object (the performer's finger in our case) is measured through a change in capacitance of the electrode to ground. These sensors provide information relative to the presence of fingers into that specific fret. Moreover, depending on the number of fingers present in a given fret, the position of these fingers, and the pressure of the fingers to the strings, the response of the sensors differ.

Figure 1. Gesture caption system based on capacitive sensors (mounted on the fretboard) and Arduino (mounted on the body).

Capacitive variations are collected by Arduino [1], an open-source electronics prototyping platform, programmed using Capsense [2], a capacitive sensing library for Arduino. Capsense converts the Arduino digital pins into capacitive sensors that are used to sense the electrical capacitance of the human body. The acquisition system is shown in Figure 1.

As reported in [11], capacitive sensors can be noisy, and crosstalk between measured capacitances at different frets may appear. Moreover, the finger position in a given left hand situation is never exactly the same, depending on musical parameters (loudness, style, etc.) or the player (length of the fingers, etc.). Because of these two reasons, collected data can not be directly processed, and we propose the use of automatic classification techniques to tackle the problem.

3. RECORDINGS

3.1 Score

In usual guitar playing conditions, the index finger over a fret (not necessarily pressing) defines a position. The following fingers are, by default, on the three following frets. Then, the score defines the exceptions to this default fingering. Beyond that, although the score does not modify this default fingering, the fingers can press different strings. So, we have to address our problem in two dimensions: (1) the overall hand position, defined by the index finger, and (2) the played strings at that position, from 1 (high pitch string) to 6 (low pitch string). The huge number of possible finger combinations forces us to organize them according to a given criterion. The parameters we can play with are:

Hand position: The hand can move up and down the fretboard. In our case, the number of sensorized frets is 10, which allows us to move the hand from fret 1 (with fingers over frets 1, 2, 3, and 4) to fret 7 (with fingers over frets 7, 8, 9 and 10), using the default fingering.

Finger positions: Each fret can be excited by a different number of fingers. We consider there are 5 possible situations: 0, 1, 2, 3 and 6 (bar). "0" means that the fret is not active (i.e. there is no finger acting on that fret), "1" means that only one finger is acting on that fret, whatever the string is pressing, and so on. A 6 (bar) means that the full index finger is acting on that fret all over the strings. We also made some recordings with a half-bar (pressing only strings 1, 2 and 3), but for this study, we consider half-bars as normal bars.

Pressed strings: For each finger position and default fingering, there are multiple combinations for pressing strings, as shown in Table 1. From all the available combinations, there are some which are not really used because of (a) the hand can not physically hold that combination, or (b) they have no musical meaning. The highlighted combinations in Table 1 represent the recorded cases.

1st. finger	1 finger/fret	2 fingers/fret	3 fingers/fret
bar	**6000**	6200	6300
	6001	6201	6030
	6010	**6210**	6003
	6011	**6020**	
	6100	6021	
	6101	**6120**	
	6110		
	6111		
1 finger	**1000**	**1200**	1300
	1001	1201	1030
	1010	1210	1003
	1011	1020	
	1100	1021	
	1101	1120	
	1110		
	1111		
2 fingers		**2000**	
		2100	
		2200	

Table 1. Finger activation combinations for each default fingering position. 4 digits refer to 4 successive frets. Each digit corresponds to the number of fingers pressing at the same fret. These positions can be played in different hand positions and in different strings. *6* refers to bar activation, *1* refers to 1 finger activation at any string, *2* refers to 2 finger activation at the same fret at any strings, and *3* refers to 3 finger activation at the same fret at any strings. The highlighted combinations represent the recorded cases.

Beyond that, it is important to distinguish between positions that can seem similar (i.e. *1000* and *0010*) but the hand position is completely different and, as a consequence of that, the residual capacitive measure from the other fingers is different. The use of one of these two options is determined by the musical context, which is not covered in this paper. Then, for simplicity, we will skip these alternative recordings.

[1] www.arduino.cc
[2] http://www.arduino.cc/playground/Main/CapSense

Position	Played strings	Category
1000	s1, s2, s3, s4, s5, s6	1000a
1010	s5s6, s4s5, s3s4, s2s3, s1s2	1010a
1010	s4s6, s3s5, s2s4, s1s3	1010b
1010	s3s6, s2s5, s1s4	1010c
1100	s5s6, s4s5, s3s4, s2s3, s1s2	1100a
1100	s4s6, s3s5, s2s4, s1s3	1100b
1100	s3s6, s2s5, s1s4	1100c
1110	s5s4s6, s4s3s5, s3s2s4, s2s1s3	1110a
1110	s4s5s6, s3s4s5, s2s3s4, s1s2s3	1110b
1200	s5s6s4, s4s5s3, s3s4s2, s2s3s1	1200a
1200	s4s6s5, s3s5s4, s2s4s3, s1s3s2	1200b
2000	s6s5, s5s4, s4s3, s3s2, s2s1	2000a
2100	s6s4s5, s5s3s4, s4s2s3, s3s1s2	2100a
2100	s5s4s6, s4s3s5, s3s2s4, s2s1s3	2100b
2200	s5s3s4s2, s4s2s3s1	2200a
6000	full, half	6000a
6010	s5, s2	6010a
6020	s5s4, s4s3	6020a
6100	s5, s2	6100a
6110	s3s5, s2s4	6110a
6120	s3s5s4, s2s4s3	6120a
6210	s4s3s5, s3s2s4	6210a

Table 2. Detailed list of all the recorded positions, specifying the played strings. Each recording includes the hand position moving from fret 1 to 7. The s1..s6 stands for the played string. Each string specification follows an ascending order from finger 1 to 4. In this paper, we refer these positions according to the *Category* column.

The recorded positions are not all the complete combinations. The recorded subset represents, under our point of view, the most common situations in real guitar performances, and also covers some specific situations in which the position recognition presents a difficulty (i.e. *6020* vs *6120*). For each of the proposed positions, several string combinations have been recorded. The same configuration of fingers over the frets also include different possibilities. For instance, the position *1200* may represent *Am* with fingers 1,2, and 3 at strings 2, 4 and 3, respectively, or the *Emaj* with the same fingers at strings 3, 5, and 4, respectively, by moving the whole hand 1 string down. In our analysis, we consider these positions are equivalent. Beyond that, the same position *1200* may represent *Am* with fingers 1,2, and 3 at strings 2, 4 and 3, respectively, or *D7* with fingers 1, 2, and 3 at strings 2, 3 and 1, respectively. Note how the order of the fingers has changed. In our analysis, we study whether these positions present an equivalent response or not.

Table 2 shows a detailed list of all the recorded positions, specifying the played strings. Each recording includes the hand position moving from fret 1 to 7. From the multiple options for each configuration, we have used that one covering the worst case, i.e., we recorded *6110s3s5* instead of *6110s5s3* because, in the first case, the hand is near the fretboard producing a higher crosstalk between the

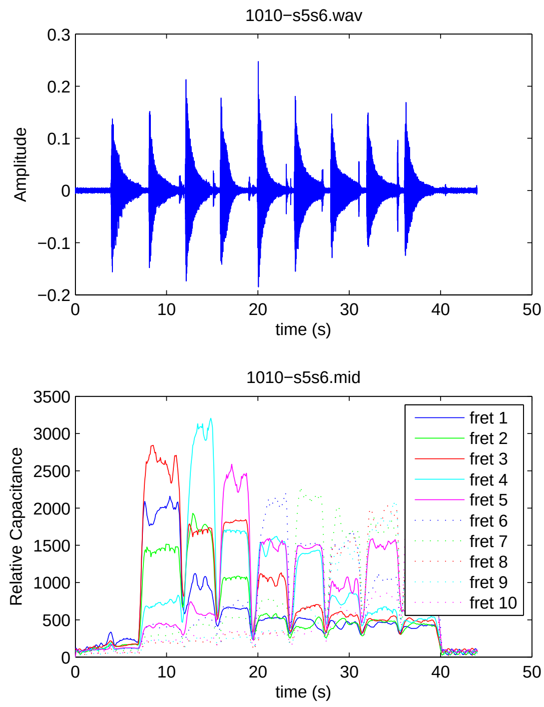

Figure 2. recorded audio and data from capacitive sensors for the 1010-s6s5 position.

measured data from frets.

3.2 Data processing

For all the recordings detailed in Table 2, the audio data from a microphone, and the data from capacitive sensors is captured. Data from capacitive sensors is converted to MIDI. We use 10 MIDI channels, one for each fret, and the information is stored as PitchBend messages to obtain a better resolution. As explained in [11], MIDI data provided by the Arduino does not have a constant sampling rate. We apply automatic resampling obtaining a constant sampling rate sr=30[Hz], which is quite low but accomplishes our requirements. Each hand position has a duration of 4 beats in a 4/4 bar at 60[bpm]. The first bar is used as pre-roll, the second bar is used to play an open strings position in all the recordings, and the specified position starts at the 3rd. bar. Figure 2 shows an example of the recorded audio and gestural information. All the recorded MIDI files can be downloaded at *www.iiia.csic.es/guitarLab/*.

The goal of this paper is to obtain models for each fingering position. We assume the collected data for the same position played at different hand positions is similar (that is, from frets 1 to 7). Then, we collapse all the information for each recording (moving the hand from fret 1 to 7) and build a pattern for that finger position. In order to avoid possible variations produced by the hand movements, we only use information from beats 2 and 3 of every bar, in which we assume the hand position is stable, and compute the mean for all the acquired data from sensors in this period of time. We know the extracted information from bars

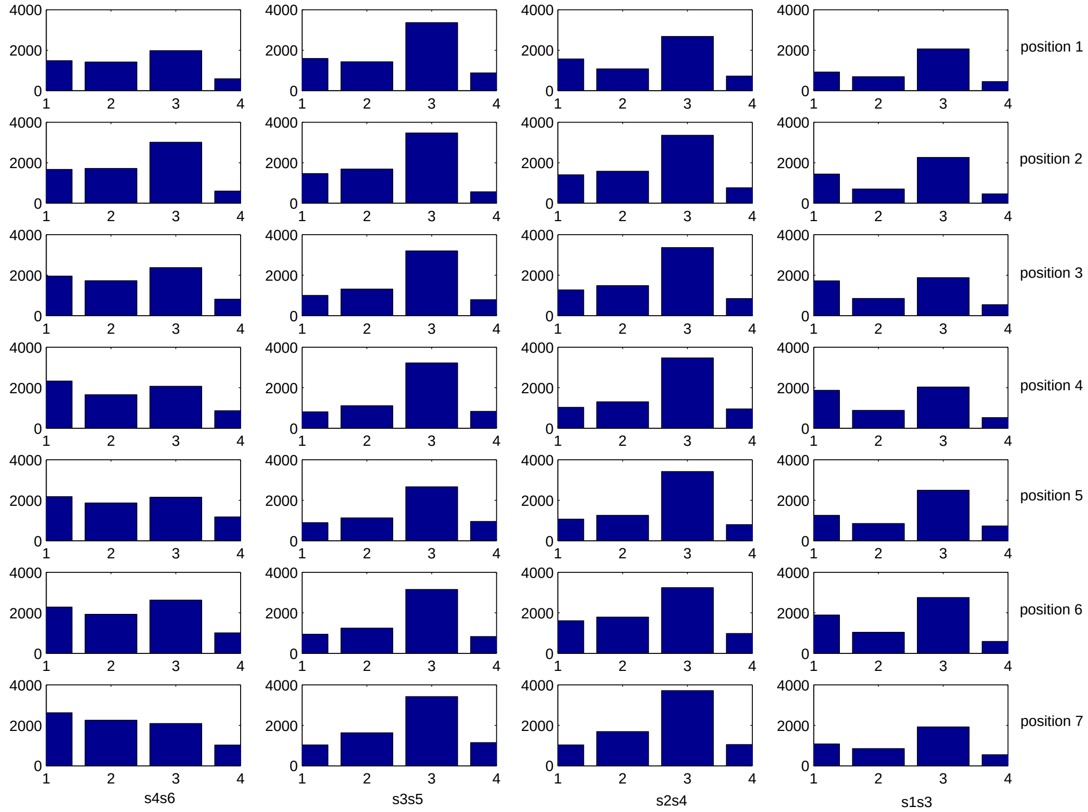

Figure 3. Patterns for finger positions 1, 2, 3 and 4 with respect to the fret position of the index finger, collected for the *1010b* position. Each column corresponds to the same pattern modifying reference frets (from 1 to 7), and each row corresponds to the same pattern modifying strings (s4s6, s3s5, s2s4, and s1s3). Vertical scale refers to measured capacitance.

2 and 3 may differ from the information obtained in other scores (we are a bit conservative, here) but our goal is to obtain the patterns in which the real and faster recordings will be compared to. These means are used to build a pattern for finger positions 1, 2, 3 and 4 with respect to the fret position of the index finger. After some preliminary experiments, we may assume that the information from the other frets is not relevant. These patterns are also collapsed through playing the same finger position at different strings.

In summary, we create, for each position, a pattern for frets 1 to 4 (relative to the position of the index finger) moving the position horizontally on the fretboard (moving the hand from low pitches to high pitches) and vertically (moving the strings from low pitch to high pitch). Figure 3 shows an example of some individual patterns collected to create the *1010* position. Plottings for all the patterns can be downloaded at *www.iiia.csic.es/guitarLab/*.

4. ANALYSIS

In this section, we present the patterns that define different finger positions and the analysis of collected data. Specifically, we are interested in verifying the following hypotheses: (H1) Moving up and down the same position through strings does not change the pattern; (H2) Moving up and down the same position through the fretboard does not change the pattern; (H3) The presence of a bar is always detected and it does not mask the information of following frets; (H4) Positions with one finger per fret can be detected; (H5) Positions with more than one finger per fret can be detected; and (H6) Different finger positions under the same fret configuration present a different the pattern.

The analysis of the collected data is divided in three parts. First, we describe how patterns are created. Then, we analyze whether the obtained patterns are coherent with what we expected. Finally, we analyze whether the obtained patterns can discriminate between different positions automatically.

4.1 Pattern creation

For all the obtained patterns (some of them are shown in Figure 3) and for each recorded position, the behavior is similar. This means that the given values and slopes are equivalent for each row, that is, the same pattern is obtained by playing at different reference frets by moving the hand horizontally on the fretboard, and for each column,

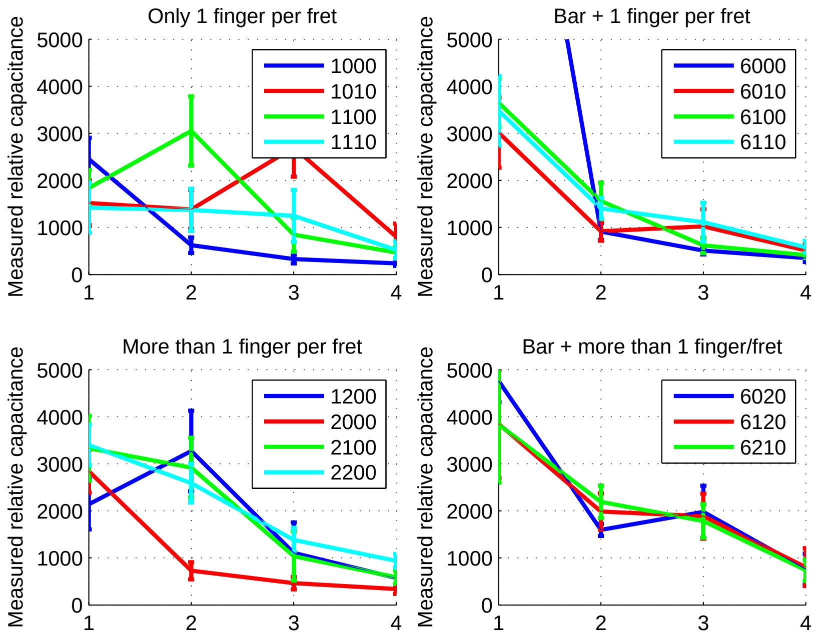

Figure 4. Patterns obtained from means and standard deviations for all the recordings at different finger positions. Position *6000* jumps to 16000, but we skip the vertical scale from 0 to 5000 to ease visual comparison.

that is, the same pattern is obtained by playing at different strings. This result verifies the hypotheses H1 and H2. Then, we may group the patterns for each detailed position in Table 2 into one of the the 22 proposed categories.

4.2 Study of patterns

As expected, the patterns captured with capacitative sensors are not linear combinations of the basic *1000*, *0100*, *0010*, and *0001* patterns. That is, the finger positions are influencing neighboring frets. However, the slopes are consistent with the activated frets. For instance, *6210* recordings present a descending slope whereas *6120* recordings tend to emphasize a sub-peak at third fret.

Regarding finger combinations with a bar, the experiments demonstrate that the presence of a bar does not mask the other fingers (see Figure 4). Indeed, the presence of a bar generates more stable positions (diminishing standard deviation). This result verifies hypothesis H3. Positions *1000* and *2000* can be confused because the slope is similar and the unique difference is the absolute value of the first fret. Although this value is higher at position *2000*, the difference is not large enough to establish a decision point.

Finger combinations in which consecutive frets are activated, present a more clear behavior, both in terms of slope and small deviation. The clearest exponents are recordings with only one finger (*1000*) or a bar (*6000*) pressing the strings, but positions like *1110*, *1200*, *2000*, *2100*, and *2200* follow also a clear behavior.

Two finger combinations require a deeper analysis: *1100* and *1010* (see Figure 4). The two combinations were played with the second active finger pressing lower strings, and lower capacitive values were expected. But higher values were obtained. Regarding position *1100*, the measured relative capacitance is really similar to position *1200*, thus, our system won't be able to distinguish among these to finger combinations. Regarding position *1010*, the first finger sometimes causes a low activation (see Figure 3). Moreover, because the middle finger tends to be close to the fretboard, the measured relative capacitance in the second fret is similar to the measured when one finger is present. These observations partly verify hypotheses H4 and H5, and the use of an automatic classification algorithm will help us to study them in detail.

4.3 Automatic detection

Once we have verified the measured patterns mostly agree with the expected ones, we analyzed whether an automatic classifier might identify them. We have 22 categories (including 75 possible finger combinations) recorded at 7 reference fret positions, that is, a data-set with 525 recordings. As discussed in Section 4.2, not only the absolute values are important in the analysis, but the slopes. In order to include slope relative information to the system, we computed the difference of the means from one fret with respect to the previous one.

The baseline for random classification is 1/22=4,54%. For simplicity, we use a K-nearest neighbours classifier (with K=3) and evaluate using 10-fold cross validation. Results provide an overall accuracy of 44,6% (weighted averaged precision = 0.449, weighted averaged recall = 0.446, weighted averaged f-measure = 0.435). The confusion matrix is shown in Figure 5, and precision and recall values for individual categories are shown in Table 3.

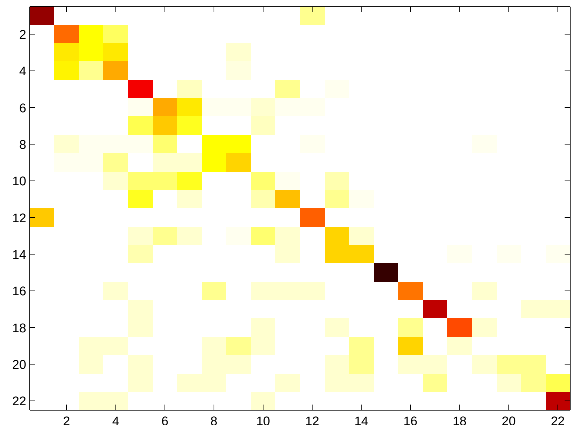

Figure 5. Results of automatic classification using K-nearest neighbours with K=3. Rows indicate categories that should be classified and columns indicate automatically classified categories. Indexes follow these categories: (1)1000a, (2) 1010a, (3) 1010b, (4) 1010c, (5) 1100a, (6) 1100b, (7) 1100c, (8) 1110a, (9) 1110b, (10) 1200a, (11) 1200b, (12) 2000a, (13) 2100a, (14) 2100a, (15) 2200a, (16) 6000a, (17) 6010a, (18) 6020a, (19) 6100a, (20) 6110a, (21) 6120a, and (22) 6210a.

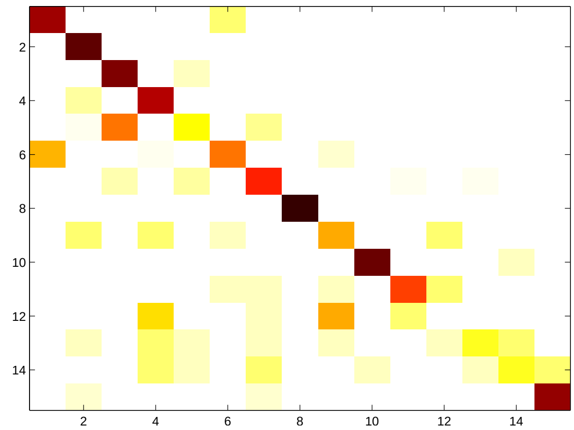

Figure 6. Results of automatic for collapsed categories classification using K-nearest neighbours with K=3. Rows indicate categories that should be classified and columns indicate automatically classified categories. Indexes follow these categories: (1)1000, (2) 1010, (3) 1100, (4) 1110, (5) 1200, (6) 2000, (7) 2100, (8) 2200, (9) 6000, (10) 6010, (11) 6020, (12) 6100, (13) 6110, (14) 6120, and (15) 6210.

Rows indicate categories that should be classified and columns indicate automatically classified categories

Position 16 (*6000*) is perfectly classified. But we observe some confusions between the other positions. First, indexes 1 and 12 (positions *1000* and *2000*, respectively) are the worst classified, but confusions are only among them. This is an expected confusion and it does not affect the identification from the different positions at all. Beyond that, we also observe important confusions between indexes 2, 3, and 4, which correspond to positions 1010a, 1010b, and 1010c, respectively. Note how all these confusions belong to position *1010*, but changing the finger's order, that is, they are equivalent. The number of fingers pressing the frets is the same, and our sensor is not designed to distinguish between them. In a similar way, more confusions can be found between indexes 5, 6, and 7 (positions *1100a*, *1100b*, and *1100c*, respectively), between indexes 8 and 9 (positions *1110a* and *1110b*, respectively), between indexes 10 and 11 (positions *1200a* and *1200b*, respectively), and between indexes 13 and 14 (positions *2100a* and *2100b* respectively). But the number of fingers over the frets is always the same. In the forthcoming positions, with the presence of a bar, confusions are more spread in the space, because the number of fingers on the frets is maximum.

We repeat the automatic classification process by collapsing the equivalent positions (see Table 2). With the resulting 15 categories, and the baseline for random classification is 1/15=6.67%, We achieved an overall accuracy of 69.5% (weighted averaged precision = 0.67, weighted averaged recall = 0.695, weighted averaged f-measure = 0.673). The confusion matrix is shown in Figure 6, and precision and recall values for individual categories are shown in Table 3. Only significant two confusions are still remaining: between positions *1000* and *2000*, and between positions *1100* and 1200, as reported in Section 4.2. For the other finger combinations, confusions are not significative and more spread in the space. Thus, the behavior of the automatic classifier is coherent. To conclude, hypotheses H4 and H5 are partially verified, and hypothesis H6 is verified.

5. CONCLUSIONS

The overall goal of our research is to understand expressivity in guitar performances through particular articulations used by different players, styles or musical genres. For that, we need to capture gesture information from the left hand to analyze the fingering and possible articulations.In this context, this paper presented a model that detects the left hand position, based on gesture information, using classification techniques.

We proposed an acquisition system based on capacitive sensors, we discussed the scores and formats for recordings and analyzed the results directly from the data and using a state of the art automatic classifier. We proposed a list of hypotheses that were practically verified, but results using the proposed automatic classifier can be improved. For that, more research is required. Specifically, we will focus our efforts on improving the gesture acquisition system, by including information from hexaphonic pickup, and musical context information to the classification algorithm.

6. 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.

Position	Expanded Precision	Expanded Recall	Collapsed Precision	Collapsed Recall
1000 a	0.735	0.857	0.714	0.833
1010a	0.324	0.524	0.85	0.944
1010 b	0.333	0.286		
1010 c	0.395	0.429		
1100 a	0.333	0.714	0.681	0.889
1100 b	0.286	0.429		
1100 c	0.250	0.257		
1110 a	0.364	0.286	0.722	0.813
1110b	0.357	0.357		
1200a	0.227	0.179	0.483	0.292
1200b	0.478	0.393		
2000a	0.655	0.543	0.625	0.500
2100a	0.323	0.357	0.689	0.646
2100 b	0.556	0.357		
2200a	0.688	0.786	0.750	0.857
6000 a	1.000	1.000	1.000	1.000
6010 a	0.438	0.50	0.333	0.417
6020 a	0.786	0.786	0.917	0.917
6100 a	0.727	0.571	0.636	0.583
6110 a	0.000	0.000	0.00	0.000
6120a	0.500	0.143	0.500	0.250
6210 a	0.400	0.143	0.500	0.250

Table 3. Precision and recall for automatic classification for (a) all the fingering positions individually classified (See Figure 5), and (b) collapsed fingering positions (See Figure 6).

7. REFERENCES

[1] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*. Springer-Verlag, New York, NY, 1988.

[2] B. E. Richardson, "Classical guitar construction: The acoustician's tale," *The Journal of the Acoustical Society of America*, vol. 117, p. 2589, April 2005.

[3] 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.

[4] D. Young, "The hyperbow controller: Real-time dynamics measurement of violin performance," in *Proc. of New Interfaces for Musical Expression*, 2002.

[5] M. Laurson, C. Erkut, V. Välimäki, and M. Kuuskankare, "Methods for modeling realistic playing in acoustic guitar synthesis," *Comput. Music J.*, vol. 25, no. 3, pp. 38–49, 2001.

[6] M. M. Wanderley and P. Depalle, "Gestural control of sound synthesis," *Proc. IEEE*, vol. 92, pp. 632–644, April 2004.

[7] A. Burns and M. Wanderley, "Computer vision method for guitarist fingering retrieval," in *SMC'06: Proceedings of the Sound and Music Computing*, May 2006.

[8] A. Burns and M. Wanderley, "Visual methods for the retrieval of guitarist fingering," in *NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression*, pp. 196–199, June 2006.

[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] 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.

[11] E. Guaus, T. Ozaslan, E. Palacios, and J. Arcos, "A left hand gesture caption system for guitar based on capacitive sensors," in *Proceedings of NIME-2010*, pp. 238–243, 2010.

[12] J. Paradiso and N. Gershenfeld, "Musical applications of electric field sensing," *Computer Music Journal*, vol. 21, no. 2, pp. 69–89, 1997.

[13] S. Hughes, C. Cannon, and S. Modhráin, "Epipe : A novel electronic woodwind controller," in *Proc. of New Interfaces for Musical Expression*, pp. 199–200, 2004.

[14] E. Miranda and M. Wanderley, *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard (Computer Music and Digital Audio Series)*. A-R Editions, Inc., 1st ed., 2006.