A Machine Learning Approach to Violin Bow Technique Classification: a Comparison Between IMU and MOCAP systems

David Dalmazzo
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
david.cabrera@upf.edu

Simone Tassani
SIMBIOsys (DTIC)
Universitat Pompeu Fabra
Barcelona, Spain
simone.tassani@upf.edu

Rafael Ramírez
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
rafael.ramirez@upf.edu

ABSTRACT

Motion Capture (MOCAP) Systems have been used to analyze body motion and postures in biomedicine, sports, rehabilitation, and music. With the aim to compare the precision of low-cost devices for motion tracking (e.g. *Myo*) with the precision of MOCAP systems in the context of music performance, we recorded MOCAP and *Myo* data of a top professional violinist executing four fundamental bowing techniques (i.e. *Détaché*, *Martelé*, *Spiccato* and *Ricochet*). Using the recorded data we applied machine learning techniques to train models to classify the four bowing techniques. Despite intrinsic differences between the MOCAP and low-cost data, the *Myo*-based classifier resulted in slightly higher accuracy than the MOCAP-based classifier. This result shows that it is possible to develop music-gesture learning applications based on low-cost technology which can be used in home environments for self-learning practitioners.

CCS CONCEPTS

• Human-centered computing → Gestural input;

KEYWORDS

Gesture, Machine Learning, MOCAP, *Myo* Armband, Audio Descriptors

ACM Reference Format:

David Dalmazzo, Simone Tassani, and Rafael Ramírez. 2018. A Machine Learning Approach to Violin Bow Technique Classification: a Comparison Between IMU and MOCAP systems. In *5th international Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '18)*, September 20–21, 2018, Berlin, Germany. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3266157.3266216>

1 INTRODUCTION

Chomsky (1975) on his book *Powers and Prospects: Reflections on human nature and the social order*, argues that language appeared in the homo-sapiens about 100,000 years ago, as a precise sonification form evolved from a gestural-based communication system [10]. Furthermore, *gestures* can be defined as a form of non-verbal

communication associated with an intention or an emotional state articulation. However, they are not only intrinsically part of the human language, but also explain specific details of a body-knowledge execution or performance. Armstrong et al. (1995) argue that gestures have an underlying brain mechanism common in both language and motor functions [1]. From this perspective, gestures are being studied not only in the language research but also in dance, performance, sports, rehabilitation, and music; the term gesture is understood as a "*learned technique of the body*" [8]. In the music context, musical phrases can be directly correlated to gestures and motor-abilities, for instance, melody definition and execution can be understood as an extension of a discrete gesture which has three states, starting, a form (shape) and its ending.

In highly competitive sports, in music education and professional context, gestures are assumed as automatic-motor abilities, learned by repetitional practice, to self-teach and fine-tune the body to execute an action optimally. Therefore, those gestures are intended to be part of the technical performer repertoire to take fast decisions in on-the-flight actions. Gestures are not only relevant in music expressive capabilities, also, a method for a correct "*energy-consumption*" habit development to avoid injuries.

Current motion capture technologies are capable of detecting body motions details very accurately, and they have been used in a variety of sports industries to enhance athletes throughput, or in rehabilitation applications [9]. Bilodeau et al. (1959) argue that real-time feedback has a more positive effect on the self-learning practice acquiring new motor skills [3]. Furthermore, in the musical education context, tracking systems and inertial motion units (IMU) are being developed recently with the aim to implement similar computer-assisted methodologies to improve educational instruction and performance. Motion Capture Systems (MOCAPS) based on 3D tracking dots by infrared cameras in a prepared room or electromagnetic positional tracking systems such as *Polhemus* can be quite expensive. Hence, new models using wearable devices based on magnetometers, gyroscopes and accelerometers, in conjunction with machine learning algorithms are being reported as efficient and low-cost solutions for analyzing body motion and gestural information.

1.1 Motivation

The framework where this study is being developed is the European project TELMI (*Technology Enhanced Learning of Musical Instrument Performance*), which investigates how technology (computer systems, sensors and software) can help music students to focus on precise information with the aim to develop better musical skills [21]. The project is a collaboration between the Universitat Pompeu

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

iWOAR '18, September 20–21, 2018, Berlin, Germany

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6487-4/18/09...\$15.00

<https://doi.org/10.1145/3266157.3266216>

Fabra, the University of Genova and the Royal College of Music, London. With the focus on violin performance, one main goal of the project is to provide real-time feedback to students about their performance relative to a good-practice models. In a previous work, we implemented a Hidden Markov Model to provide real-time feedback to self-practice learners; in this study, we compare two motion capture methods - the expensive and highly precise MOCAP system and the low-cost Myo device [22] - in order to determine if the motion data provided by both devices affect the accuracy of the corresponding machine learning models for classifying bow strokes in violin performance.

2 RELATED WORK

Sawada et al. (1997) reported the configuration of an IMU device made with accelerometer sensor. They describe rotational force and directional attributes extracted from the sensors, to classify gestural expressions in musical context. Their motivation is to measure non-verbal communication and emotional intentions in music performance. They applied tempo recognition in orchestra conductors to describe how the gestural information is imprinted in the musical outcome [18].

Peiper et al. (2003) presented a study of violin bow articulations classification. Their method describes a decision tree (ID3) algorithm which is modelled to identify four standard bow articulations, *Détaché*, *Martelé*, *Spiccato* and *Staccato*. The gestural information is extracted using an electromagnetic motion tracking device mounted close to the performer right hand. The visual outcome is displayed in a CAVE as a room with a four-wall projection setup for immersive virtual reality applications and research. Their system reported high accuracy (around 85%) when tested against two gestures, however, decrease to 71% accuracy, comparing more than four articulations [17]. Gibet et al. (2006) developed an "augmented violin" as an acoustic instrument with aggregated gestural electronic-sound manipulation. They modelled a kNN algorithm for the classification of three standard violin bow strokes known as *Détaché*, *Martelé* and *Spiccato*. Authors describe how an Analog Device ADXL202 is placed at the bow-frog to transmit inertial motion information. It uses two accelerometers which report bowing direction in contraposition of the violin strings. The custom hardware developed at IRCAM, send through a wireless RF receiver the IMU's data to a Max/MSP Ethernet connection using OSC protocol. Gibet's study gives the background for next studies placing wireless electronics in classical instruments. He describes a liner discrete analysis to identify important spacial dissimilarities among bow articulations, giving a highly accurate gestural prediction in the three models presented (*Detaché* 96.7%, *Martelé* 85.8% and *Spiccato* 89.0%). On the other hand, they described a k-NN model with 100% accuracy estimation in *Detaché* and *Martelé*, 68.7% in *Spiccato*. They conclude that accuracy is directly linked to dynamics, fluctuating its predictions depending on dynamical emphasis ranges within *pp*, *mf* and *ff* [14]. Caramiaux et al. (2009), has presented a real-time gesture follower and recognition model based on HMM. The system gives an output of continuous numerical descriptors related to a time progression and its likelihood; its application is described for different contexts such as music education, music and dance performances and interactive installations [5]. Our research is directly

linked to this previous study, where we implemented the Max 7 version of this model, included in the MUBU citeschnell2009mubu externals developed at IRCAM.

Vatavu et al. (2009) have proposed a Naive-Detection algorithm to discretize temporal event in a two-dimensional gestural drawing matrix. It describes the similitude between two gestures (template vs new drawing) with a minimum alignment cost between the curvature functions of both gestures [24]. Bianco et al. (2009) address the question of how to describe acoustic sound variations directly mapped to its unique gesture performance within its articulations based on PCA and segmentation in their sound analysis. They have focused on the study of a professional trumpet player, performing a set of exercises with specific dynamical changes [2]. In this study, it is identified that the description of the relationship between gestures and sound is not linear, hypothesizing the at least two motor-cortex control events are involved in the performance of single notes.

Zamborlin et al. (2010) presented a study conducted at IRCAM where an HMM-based system is implemented with specific detail of the Markov chain setup for temporal segmentation. The authors define a model where discrete segmentation of gestural data is limited to a maximum of 600 states in a minute range phrase where upper segmentation might be too CPU demanding for nowadays personal computers. Their goal is not to describe a specific gestural repertoire, instead, propose an optimal "low-cost" algorithm for any gestural classification without the need for big datasets [25]. Zamborlin method is then chosen as the machine learning techniques applied in this study developed at MTG. Gillian et al. (2011) exposed a different approach than the standard Markov Model previously described. On their research, a Dynamic Time Warping (DTW) is extended to classify N-dimensional signal with a low number of training samples, having an accuracy rate of 99%. To test DTW algorithms, the authors first defined a set of 10 gestures as an "air drawing" articulations of the right hand. Their methodology gives a valid and optimal approach to classify any gesture [15].

A study conducted by VanDerLinden et al. (2011) described the invention of a set of sensors and wearables called *MusicJacket*. They aim to give postural feedback and bowing technique references to novice violin players. Authors reported that vibrotactile feedback directly engages the subjects' motor learning systems, correcting their postures almost immediately, shortening the period needed to acquire motor skills and reduces cognitive overload [23]. Schedel et al. (2011), have implemented the Wekinator application [11] to classify seven standard cello bow articulations such as *legato*, *spiccato*, or *marcato*, among others; using a commercial IMU device known as K-Bow for the motion data acquisition. The cello performer used a foot pedal to stop and start articulation training examples. For each stroke, she varied the string, bow position, bow pressure, and bow speed. After training a model, the cellist evaluated it by demonstrating different articulations. The authors created an interactive system for composition and sound manipulation in real-time based on the bow gesture classifications [19].

In the field of mapping gestural data to sound design in real-time, based on the "mapping by demonstration" principle exposed by Françoise [13], were the users create their gestural repertoire by simple-direct examples on the flight. Caramiaux et al. (2013) proposed a framework revising and compared different Machine

Learning algorithms to apply regression, classification, segmentation, clustering and prediction of a set of pre-established gestures. Their objectives were focused on musical gesture analysis and gestural sound generation and control [7].

Françoise et al. (2014) presented a set of probabilistic models included in a Max external (MUBU [20]) and compared their features for real-time sound mapping manipulation. Those algorithms are Gaussian Mixture Models (GMM), Gaussian Mixture Regression (GMR), Hierarchical HMM (HHMM) and Multimodal Hierarchical HMM (MHMM). Giving as a conclusion a detailed explanation on which one is more suitable depending on the nature of the sound interaction; defining a hierarchy on movement models, multi-modal models, instantaneous and temporal models [12][13]. Caramiaux et al. (2014) have proposed a different approach of gesture recognition using a Sequential Monte Carlo inference technique instead of the Dynamic Time Warping method. They applied the models evaluating standard 2D gestures drawn using an on-screen interactive pen [6].

Kerber et al. (2017) implemented a custom Python program to read, filter and process electromyogram data extracted from the commercially available armband device called *Myo*. Authors trained a machine learning model based on SVM, to classify and accurately (95%) detect 40 different gestures. Those gestures are based on the *Myo* standard hand dispositions, also different finger configurations [16].

3 MATERIALS AND METHODS

3.1 Music Materials

For this study, a set of four gestural violin-bowing articulations were recorded as a reference for the ideal performance by expert professional violinist Madeleine Mitchell from the Royal College of Music, London. The recordings included MOCAP and *Myo* armband data and were used to train machine learning models to classify the bowing gestures. For the music score reference see the figure 1:

- **Détaché** It means *separated*; the method describes a clean, stable sound with each bowing direction, moving smoothly from one note to the next. The weight over the violin strings is even for each note performed. It is the most common bowing technique in the repertoire. The exercise was performed within two octaves ascending and descending scale in 4/4 at 70BPM, playing three eighth-triplet per note. In total 32 gesture samples recorded.
- **Martelé** The term means *hammered*; it is an extension of *Détaché* with a more distinctive attack, caused by a faster and slightly stronger initial movement to emphasize the motion starting point and it has a moment of silence at the end. Two octaves were played at 120 BPM in 4/4 played with Quarter-notes. Having 32 gestures in total.
- **Spiccato** *Spiccato* is a light bouncing of the bow against the strings. It is achieved by the physical effect of attacking the strings on a vertical (horizontal) angular approach of the bow with a controlled weight and a precise hand-wrist control. Two octaves performed at 90BPM attacking each note with three eighth-triplets. 32 gestures in total.
- **Ricochet** (Jeté) Also known as *Jeté*: It is a controlling bouncing effect played in a down-bowed gesture with a *staccato*

Bow Strokes

Figure 1: Bow-Stroke music score guide

accent and even bouncing rhythmic accents. Usually among two to six notes are produced with a controlled rhythmic bouncing; however, up to ten rebounds can be played with one gesture attack. Two octaves were played at 60 BPM in 4/4 played with Quarter-notes attacks within eighth-triplets bounces accents.

3.2 Data Acquisition

3.2.1 Motion Capture. Madeleine bow-stroke recordings were conducted in the MOCA Lab at University Pompeu Fabra. The Motion Capture room is a marker-based system equipped with a system of eight infrared cameras with 1.5 MP resolution at 250 fps and can detect maker displacements inferior to 0.1 mm in an acquisition volume of 4x4x3 m3. The motion capture system is synchronized with video cameras and two force plates, for integrated biomechanical analyses of human body motions. The system has been provided and installed by the company BTS [4]. Markers were defined as HB Head Back, HF_L Head Front Left, HF_R Head Front Right, SB_R Shoulder Back Right, SF_R Shoulder Front Right, SB_L Shoulder Back Left, SF_L Shoulder Front Left, EF_R Elbow Front Right, EB_R Elbow Back Right, EF_L Elbow Front Left, EB_L Elbow Back Left, WF_R Wrist Front Right, WB_R Wrist Back

Figure 2: A preview of the MOCAP application *SMART-Tracker* after defining the 3D model.

Right, WF_L Wrist Front Left, WB_L Wrist Back Left, HP_L Hip Left, HP_R Hip Right, KF_1_R Knee Front Right 1, KF_2_R Knee Front Right 2, KF_1_L Knee Front Left 1, KF_2_L Knee Front Left 2, F_R Foot Right, F_L Foot Left. For the violin tree markers, one in the Right Upper Bout, one in the Left Lower Bout, and one in the Scroll. Arc has a mark in the Stick and two marks (T configuration) at the Tip. (see figure 2)

3.2.2 Myo. At the same time during the MOCAP recording session, we used the *Myo* device which is composed of a set of sensors for motion estimation and a haptic feedback motor. The bracelet size is between 19 - 34 cm adjustable to the forearm circumference. It weighs 93 grams. The hardware is composed of eight medical grade stainless steel EMG sensors that report electrical muscle activity. The IMU contains three-axis gyroscope giving degrees of change in radians per second (angular velocity), three-axis accelerometer as an estimation of -8g to 8g ($1g=9.81 \text{ m/s}^2$), three-axis magnetometer giving as an output a Quaternion reference of the imaginary rotation of the *Myo* in the space. It has an ARM Cortex M4 Processor, and it reports short, medium and long haptic feedback vibration. Its communication with a computer is based on Bluetooth with an included adapter, giving a sampling rate of 200Hz (Hop-time of 5 ms).

3.2.3 Openframeworks Implementation. It was used to acquire and visualize the IMU's information in real-time. OF is an open-source platform based on C++ which has a collection of libraries to develop applications in all operating systems. Developers commonly use it in the field of interactive applications, video games, mobile apps. Addons are additional available libraries in the community, and we have developed an *addon* (ofxMyo) to receive *Myo* information and will be released as part of the publication. With it, we recorded CSV files to generate a motion database (see figure 3).

Figure 3: Openframeworks implementation: *Myo* visualizer and CSV data recorder in sync with audio files. Essentia library implementation to extract audio features to define from audio, areas where the gesture is being performed

3.3 Audio and Data Synchronization

The first part of the synchronization was made implementing *Max 7*, which is a visual programming language platform commonly used in electronic music and interactive media development and creation, suitable for quick prototyping and it allows communication with external devices. To synchronize data coming from the *Myo* device through the C++ app, camera and audio, we build a *Max* patch which sends OSC events to the *Myo* application to generate a database which contains files in CSV format, at the same time-window range reference the audio data is created within *Max*. To acquire audio we used an interface Zoom H5 linked to *Max*, recording WAV files with a sample rate of 44.100Hz/16bits. To time align the signal from the *Myo* and the MOCAP data, we apply first a sample rate unification with linear interpolation to add extra points into the *Myo* data trajectories with a sample difference of 1:4.16666. MOCAP system records at 250fps while *Myo* was recorded at 60fps. Comparing x-axis motion from the Stick marker in the bow against *Myo* Euler angle Yaw, we aligned signal-peaks, matching starting and ending points of the bow articulations. From the audio onsets, applying Essentia library for the audio descriptors, we aligned the gesture windowing to the sound, to extract the motion data from the MOCAP database. The MOCAP data was constructed using *SMARTTracker* application from the BTS company. 3D model recorded has the points in a virtual space, and they need to be placed by hand to define the correlation from the body model described before. Once the trajectories are clean and defined, the vectors are exported into a CSV file to be then uploaded in WEKA. We trained different models such as k-NN, J48 and SVM in WEKA to compare gesture classification and recognition in both cases, MOCAP data and *Myo* data.

4 RESULTS

After implementing a sort of Machine Learning algorithms in WEKA platform, J48 Decision tree algorithm has reported higher accuracy in both samples sets. In the case of MOCAP data, J48 DT reported

99.46% of correct classification instances (table) within 28.820 samples in total, including the four gestures exposed in methods. On the other hand, Myo dataset reported 99.847% correct classified instances among 28.820 samples per all gestures. Cluster distribution of angles over the space shows how each bow-stroke has a specific space range related to the articulation, as well, the trajectory that covers each of the four strings. SVM models also reported high accuracy estimations with 99.5102% for MOCAP and 98.9183% for the Myo device, the results tables (table 10 and table 7) are included in the complementary material.

Accuracy tables (table 2 and table 5) are the results from J48 Decision-tree algorithm implemented in WEKA. True Positive Rate (**TP Rate**) and **Recall** are a normalized value expressed as a percentage of correctly predicted gestures. Gestures are displayed as *class* where 1:Détaché, 2: Martelé 3: Spiccato and 4: Ricochet. **FP Rate** is the False Positive rate. **ROC Area** is known as the *probability of detection* explained as the area formed of **TPR** against **FPR**. The weighted **F-Measure** (F-M) is the average proportion of true positives depending on how many elements are in each class.

The tables (table 1 and table 4) reports the accuracy of the correct classified instances. Both datasets, MOCAP and Myo samples were aligned and Myo data was upper-sampled (explained in methods) to match the total number of instances (28820).

MOCAP Data Results

Table 1: Stratified cross-validation for MOCAP samples

Correctly Classified Instances	28665	99.4622%
Incorrectly Classified Instances	155	0.5378%
Kappa statistic	0.9928	
Mean absolute error	0.0032	
Root mean squared error	0.0506	
Relative absolute error	0.8656%	
Root relative squared error	11.7229%	
Total Number of Instances	28820	

Summarize table of the Correctly Classified Instances in the **MOCAP** case

Table 2: Accuracy By Class (MOCAP)

Class	TP Rate	FP Rate	Prec	Recall	F-M	ROC.A
1	0.995	0.002	0.993	0.995	0.994	0.998
2	0.992	0.001	0.996	0.992	0.994	0.998
3	0.992	0.002	0.991	0.992	0.991	0.997
4	0.999	0.001	0.998	0.999	0.998	0.999
W.Avg	0.995	0.002	0.995	0.995	0.995	0.998

Numbers corresponds to: 1) Détaché 2) Martelé 3) Spiccato and 4) Ricochet.

Table 3: Confusion Matrix (MOCAP)

	a	b	c	d
a	7163	2	34	1
b	15	7015	22	18
c	37	14	5949	0
d	0	11	1	8538

The letters corresponds to: a) Détaché b) Martelé c) Spiccato and d) Ricochet.

MYO Device Data Results

Table 4: Stratified cross-validation for Myo Samples

Correctly Classified Instances	28776	99.8473%
Incorrectly Classified Instances	44	0.1527%
Kappa statistic	0.998	
Mean absolute error	0.0009	
Root mean squared error	0.0275	
Relative absolute error	0.25%	
Root relative squared error	6.3724%	
Total Number of Instances	28820	

Summarizing table of the Correctly Classified Instances in the **Myo** case.

Table 5: Accuracy By Class (Myo)

Class	TP Rate	FP Rate	Prec	Recall	F-M	ROC.A
1	1.000	0.000	1.000	1.000	1.000	1.000
2	0.998	0.001	0.997	0.998	0.997	0.990
3	0.999	0.000	0.999	0.999	0.999	1.000
4	0.998	0.001	0.998	0.998	0.998	0.990
W.Avg	0.998	0.001	0.998	0.998	0.998	0.990

The numbers corresponds to: 1) Détaché 2) Martelé 3) Spiccato and 4) Ricochet.

Table 6: Confusion Matrix (Myo)

	a	b	c	d
a	7197	0	3	0
b	0	7053	0	17
c	0	0	5996	4
d	0	20	0	8530

The letters corresponds to: a) Détaché b) Martelé c) Spiccato and d) Ricochet.

(a) MOCAP Bow's Frog trajectory per gesture

(b) Myo device trajectory per gesture

Figure 4: Samples cluster visualization for a) MOCAP Bow-Frog and b) Myo device. Color map per gesture: Détaché: Cyan, Martelé: Orange, Spiccato: Dark-blue, Ricochet: Orange-red

5 CONCLUSIONS

Comparing Myo and MOCAP data (see figure 4) it is evident that Myo is based on a Quaternion reference which reposts its imaginary position in space, it does not correspond precisely with the motion of the Bow; however, after translating to Euler angles, its patterns are clearly identifiable in terms of motion information linked to the

four gestures, being sufficient for classification and regression. It is proven that applying this values in a Hidden Markov Model it is suitable as a description of the nature of the gesture in real-time [12]. After testing accuracy from both datasets, we have confirmed that complementing low-cost devices paired with audio features extraction can be sufficient as a motion-audio descriptor, at the moment to train an ML model related to gesture detection with high accuracy.

The tables 1 and 4 report the correct classified instances in a high accuracy (99.4622% and 99.8473% respectively), demonstrating that despite the fundamental differences in trajectories, both have useful information about the classified gestures. In the supplementary materials, similar results are reported applying SVM as an ML model (see tables 7 and 10).

The confusion matrix (MOCAP) (table 3) shows a slightly correlated confusion among Spiccato and Détaché, perhaps due to the absence of sound features which will separate the nature of the samples as Spiccato has some bouncing effect over the strings. As a future work, the implementation of small motion device based on accelerometer, gyroscope and magnetometer, placed in the Bow's frog, would be suitable to add more precise motion information in conjunction with the Myo device, to then train an ML model in gestural recognition.

6 SUPPLEMENTARY MATERIALS

6.1 Tables for the SVM model

The table 7, table 8 and table 9 give the WEKA results applying SVM models in the MYO dataset, with an accuracy of Correctly Classified Instances of 98.9183%. The table 10, table 11 and table 12 give the WEKA results applying SVM models in the MOCAP dataset, with an accuracy of Correctly Classified Instances of 99.5102%.

SVM model Results for MYO and MOCAP Data

Table 7: Test mode: Split 66.0% train, for the Myo Samples (SVM)

Correctly Classified Instances	9693	98.9183%
Incorrectly Classified Instances	106	1.0817%
Kappa statistic	0.9855	
Mean absolute error	0.2509	
Root mean squared error	0.3131	
Relative absolute error	67.2711%	
Root relative squared error	72.4913%	
Total Number of Instances	9799	

Summarizing table of the Correctly Classified Instances in the **Myo** case (SVM).

Table 8: Accuracy By Class (Myo) applying SVM

Class	TP Rate	FP Rate	Prec	Recall	F-M	ROC.A
1	1.000	0.000	1.000	1.000	1.000	1.000
2	0.978	0.007	0.978	0.978	0.978	0.992
3	1.000	0.000	1.000	1.000	1.000	1.000
4	0.989	0.008	0.982	0.982	0.982	0.993
W.Avg	0.998	0.004	0.989	0.989	0.989	0.996

The numbers corresponds to: 1) Détaché 2) Martelé 3) Spiccato and 4) Ricochet.

Table 9: Confusion Matrix (Myo) applying SVM

	a	b	c	d
a	2430	0	0	0
b	0	2372	0	53
c	0	0	2034	4
d	0	53	0	2857

Training time: 1.79 seconds. The letters corresponds to: a) Détaché b) Martelé c) Spiccato and d) Ricochet.

Table 10: Test mode: Split 66.0% train for the MOCAP Samples (SVM)

Correctly Classified Instances	9751	99.5102%
Incorrectly Classified Instances	48	0.4898%
Kappa statistic	0.9934	
Mean absolute error	0.2504	
Root mean squared error	0.3122	
Relative absolute error	67.1297%	
Root relative squared error	72.3019%	
Total Number of Instances	9799	

Summarizing table of the Correctly Classified Instances in the **MOCAP** case (SVM).

Table 11: Accuracy By Class (MOCAP) applying SVM

Class	TP Rate	FP Rate	Prec	Recall	F-M	ROC.A
1	1.000	0.000	1.000	1.000	1.000	1.000
2	0.998	0.006	0.983	0.998	0.990	0.997
3	1.000	0.000	1.000	1.000	1.000	1.000
4	0.986	0.001	0.998	0.986	0.992	0.997
W.Avg	0.995	0.002	0.995	0.995	0.995	0.998

The numbers corresponds to: 1) Détaché 2) Martelé 3) Spiccato and 4) Ricochet.

Table 12: Confusion Matrix (MOCAP) applying SVM

	a	b	c	d
a	2430	0	0	0
b	0	2419	0	53
c	0	0	2034	6
d	0	42	0	2867

Training time: 1.91 seconds. The letters corresponds to: a) Détaché b) Martelé c) Spiccato and d) Ricochet.

ACKNOWLEDGMENTS

This work has been partly sponsored by the Spanish TIN project TIMUL (TIN 2013-48152-C2-2-R), the European Union Horizon 2020 research and innovation programme under grant agreement No. 688269 (TELMi project), and the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502).

REFERENCES

- [1] David F Armstrong, William C Stokoe, and Sherman E Wilcox. 1995. *Gesture and the nature of language*. Cambridge University Press.
- [2] Tommaso Bianco, Vincent Freour, Nicolas Rasamimanana, Frederic Bevilacqua, and René Caussé. 2009. On gestural variation and coarticulation effects in sound control. In *Int. Gesture Work*. Springer, 134–145.
- [3] Edward A Bilodeau, Ina McD Bilodeau, and Donald A Schumsky. 1959. Some effects of introducing and withdrawing knowledge of results early and late in practice. *J. Exp. Psychol.* 58, 2 (1959), 142.

- [4] B T S Bioengineering. 2017. {BTS Bioengineering}. <http://www.btsbioengineering.com>
- [5] Baptiste Caramiaux, Frédéric Bevilacqua, and Norbert Schnell. 2009. Towards a gesture-sound cross-modal analysis. In *Int. Gesture Work*. Springer, 158–170.
- [6] Baptiste Caramiaux, Nicola Montecchio, Atau Tanaka, and Frédéric Bevilacqua. 2014. Adaptive Gesture Recognition with Variation Estimation for Interactive Systems. *ACM Trans. Interact. Intell. Syst.* 4, 4 (2014), 1–34. <https://doi.org/10.1145/2643204>
- [7] Baptiste Caramiaux and Atau Tanaka. 2013. Machine Learning of Musical Gestures. *Proc. Int. Conf. New Interfaces Music. Expr. 2013 (NIME 2013)* (2013), 513–518. <http://nime2013.kaist.ac.kr/>
- [8] Noland Carrie. 2009. Agency and embodiment: performing gestures/producing culture. (2009).
- [9] Ed H Chi, Gaetano Borriello, Guernsey Hunt, and Nigel Davies. 2005. Guest editors' introduction: Pervasive computing in sports technologies. *IEEE Pervasive Comput.* 4, 3 (2005), 22–25.
- [10] Noam Chomsky. 1975. *Reflections on language*. New York.
- [11] Rebecca Fiebrink and Perry R Cook. 2010. The Wekinator: a system for real-time, interactive machine learning in music. *Proc. Elev. Int. Soc. Music Inf. Retr. Conf. (ISMIR 2010)* 4, 3 (2010), 2005. <http://ismir2010.ismir.net/proceedings/late-breaking-demo-13.pdf?origin=publicationDetail>
- [12] Jules François, Baptiste Caramiaux, and Frédéric Bevilacqua. 2012. A hierarchical approach for the design of gesture-to-sound mappings. In *9th Sound Music Comput. Conf.* 233–240.
- [13] Jules François, Norbert Schnell, Riccardo Borghesi, Frédéric Bevilacqua, and Place Igor Stravinsky. 2014. Probabilistic Models for Designing Motion and Sound Relationships. *Proc. Int. Conf. New Interfaces Music. Expr.* (2014), 287–292.
- [14] Sylvie Gibet, Nicolas Courty, Jean-François Kamp, Nicolas Rasamimanana, Emmanuel Fléty, and Frédéric Bevilacqua. 2006. Gesture Analysis of Violin Bow Strokes. 3881 (2006), 145–155–155. http://dx.doi.org/10.1007/11678816_{_}17{_%}0Ahttp://www.springerlink.com.ezp01.library.qut.edu.au/content/h710741265086351/fulltext.pdf
- [15] Nicholas Gillian, R Benjamin Knapp, and Sile O 'modhrain. 2011. Recognition Of Multivariate Temporal Musical Gestures Using N-Dimensional Dynamic Time Warping. *Nime June* (2011), 337–342.
- [16] Frederic Kerber, Michael Puhl, and Antonio Krüger. 2017. User-Independent Real-Time Hand Gesture Recognition Based on Surface Electromyography. (2017).
- [17] Chad Peiper, David Warden, and Guy Garnett. 2003. An Interface for Real-time Classification of Articulations Produced by Violin Bowing. *Proc. Int. Conf. New Interfaces Music. Expr.* (2003), 192–196. <http://dl.acm.org/citation.cfm?id=1085760>
- [18] Hideyuki Sawada and Shuji Hashimoto. 1997. Gesture recognition using an acceleration sensor and its application to musical performance control. , 9–17 pages. [https://doi.org/10.1002/\(SICI\)1520-6440\(199705\)80:5<9::AID-ECJC2>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1520-6440(199705)80:5<9::AID-ECJC2>3.0.CO;2-J)
- [19] Margaret Schedel and Rebecca Fiebrink. 2011. a Demonstration of Bow Articulation Recognition With Wekinator and K-Bow. *Icmc 2011 August* (2011), 272–275.
- [20] Norbert Schnell, Axel Röbel, Diemo Schwarz, Geoffroy Peeters, and Riccardo Borghesi. 2009. MUBU & friends - Assembling tools for content based real-time interactive audio processing in MAX/MSP. *Proc. Int. Comput. Music Conf. Icmc* (2009), 423–426.
- [21] TELMI. [n. d.]. Technology Enhanced Learning of Musical Instrument Performance. <http://telmi.upf.edu>
- [22] ThalmicLabs. [n. d.]. Myo. <https://www.myo.com>
- [23] Janet Van Der Linden, Erwin Schoonderwaldt, Jon Bird, and Rose Johnson. 2011. MusicJacket - Combining motion capture and vibrotactile feedback to teach violin bowing. *IEEE Trans. Instrum. Meas.* 60, 1 (2011), 104–113. <https://doi.org/10.1109/TIM.2010.2065770>
- [24] Radu-Daniel Vatavu, Laurent Grisoni, and Stefan-Gheorghe Pentiu. 2009. Multiscale detection of gesture patterns in continuous motion trajectories. In *Int. Gesture Work*. Springer, 85–97.
- [25] Bruno Zamborlin, Anthony Sypniewski, Norbert Schnell, and Nicolas Rasamimanana. 2010. Continuous Realtime Gesture Following and Recognition. (2010), 73–84. <http://articles.ircam.fr/textes/Bevilacqua09b/index.pdf>