POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Exploring novel interaction strategies in live music performances based on muscle signals and deep learning.

Laurea Magistrale in Music and Acoustic Engineering - Ingegneria della Musica e dell'Acustica

Author: Davide Lionetti

Advisor: Prof. Massimiliano Zanoni

Co-advisor: Ing. Paolo Belluco

Academic year: 2022-2023

1. Introduction

The rapid advancement of wearable devices and machine learning has paved the way for novel interaction strategies in live music performances. This research explores the field of embodied engagement with music from various perspectives such as Human-Computer Interaction, New Interface of Musical Expression (NIME). The design and exploration of gestural interactions with sound and digital media are essential in artistic practices where the performer's body engages with the music through motion and physiological sensing. This study focuses on applying a custom wearable computer to investigate the concepts of intention in relation to surface Electromyography (sEMG) for Digital Musical Instruments (DMIs) application [3]. sEMG, a voltage-based representation of muscle electrical activity, is employed to capture and sense musical gestures. sEMG allows non-invasive detection of muscle activity. The study aims to evaluate the impact of full-body muscle selection in gestural interaction design from an artistic perspective, particularly focusing on arms muscle classification relevant to guitar playing. The study fills a gap in the literature by proposing a novel interaction strategy modulating guitar effects using the musician's muscle signal, which has not been attempted before. The sEMG signals is analyzed by two deep learning models based on Bidirectional Long Short Time Memory (BLSTM) recurrent neural network [2]. This system adapts and customizes the sound based on the musician's muscle activation, expanding the compositional tools palette and fostering creativity and exploration. Finally, to evaluate the quality of the system, a questionnaire has been developed and submitted to an advanced guitarist.

2. Methods and materials

2.1. Surface Electromiography

The sEMG technique involves placing electrodes on the skin to detect the electrical activity produced by skeletal muscles. Unlike invasive EMG investigations that use needles, sEMG allows non-invasive detection of muscle activity.
The sEMG signal is a composition of the electric activity from the nearest muscular tissue, which is closely related to muscle contraction.

It consists of two distinct states: the rest state, where muscle fibers exhibit an electric potential of approximately -80 mV, and the contraction state, where muscle fibers generate electric potentials within motor units (MUs). Motor unit action potentials (MUAPs) are formed when motor neurons trigger neuromuscular junctions, resulting in intracellular action potentials propagating through depolarization and repolarization of muscle fibers. During muscle contraction, the EMG signal represents a linear combination of multiple MUAP trains. There are two types of muscle contractions: static and dynamic. Static contractions involve muscle fibers contracting without any change in length or joint motion, while dynamic contractions involve changes in muscle fiber length and joint motion. To sense the sEMG signal, at least two electrodes are required for measuring differential voltage, along with a third electrode as a DC reference. The sEMG signal ranges from -5mV to 5mV. Electrodes are designed to optimize electrode-skin impedance and minimize crosstalk from adjacent units. Due to the stochastic, nonlinear, and nonstationary nature of the sEMG signal, it is impractical to analyze the raw signals directly. EMG signals are noisy, with a lower Signal to Noise Ratio (SNR) compared to other biosignals. The noise is caused by equipment, electromagnetic radiation, motion artifacts, and crosstalk from neighboring muscles. Preprocessing is necessary to filter out unwanted noise, especially in multiclass classification problems.

2.2. Data acquisition

To integrate sEMG signals into artistic practices reliably and robustly, the LWT3 (*www.lwt3.com*) non-invasive wearable device was used. This wearable computer system enables the tracking of biometric signals from multiple muscle areas with a sampling rate of 1000Hz and 22-bit resolution ADC for 8 channels in a double differential configuration. The acquired data is then transmitted to the platform analysis via a wired USB 2.0 communication channel. The placement of the medical electrode pads has been made by following the protocols of the *Atlas of Muscle Innervation Zone* [1], in order to minimize cross-talk and to enhance the acquisition quality.
The data elaboration platform (named Raw Power) supports the acquisition of sEMG signals from the wearable device facilitating real-time visualization of muscular activities through dynamic plots. To address movement artifacts and electromagnetic interferences during live performances, a pre-filtering stage is implemented. This stage includes a fifth-order Butterworth bandpass filter (30-300 Hz) to attenuate high-frequency motion artifacts and a harmonic band-stop Notch filter (centered at 50 Hz) to suppress power line noise.

2.3. Feature And Muscle Selection

A feature selection stage is performed to enhance the discrimination capabilities of sEMG signals for various guitar techniques, The focus is on identifying two significant low-level features in the time domain:

1. **Root Mean Square (RMS)**, it is related to the constant force and non-fatiguing contraction,it reflects power activation and is directly proportional to the exerted force. It relates to standard deviation, which can be expressed as

$$RMS = \sqrt{\frac{\sum_{k=1}^{N} x_k^2}{N}}. \tag{1}$$

where N denotes the length of the signal and x_n represents the EMG signal in a segment n.

2. **Zero crossing (ZCR)** is the number of times that EMG signals crosses zero in a window of lentgh N. The threshold value is 20 mV. It can be formulated as:

$$ZCR = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} \begin{cases} 1 & \text{if } x(i) \cdot x(i+1) < 0 \\ & \text{and } |x(i) - x(i+1)| > \text{thr.} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

ZCR is related to slope sign change (SSC) and it gives a rough estimation in the frequency domain.

The muscle selection process consists of two stages. In the first stage, the average ZCR values are analyzed to assess the extent of muscle activation during the gestures. The top eight muscles are identified based on their average ZCR values. In the second stage, the average RMS values are evaluated for the selected muscles to determine the amplitude threshold.

Through extensive testing, we observed that in a resting position (with our configuration test), the ZCR values ranged from 40 to 70, while the RMS around 2.3 mV compared to the baseline. These observed ranges serve to define an heuristic threshold for muscle classification: ZCR 50 slop change and 4.6 mV into a 500 ms window. The muscles are ranked based on their mean ZCR and RMS values, discarding those below the thresholds, resulting in the selection of the most relevant and activated muscles for accurate gesture discrimination.

2.4. Selected Muscles

We propose a systematic evaluation experiment to select muscles relevant to guitar performance using surface electromyography (sEMG) data. Twelve upper limb muscles were analyzed *[Left Flexor Carpi Radiali, Left Extensor Carpi Radialis, Left Bicep Brachii Short Head, Left Brachioradialis, Right Flexor Carpi Radialis, Right Extensor Carpi Radialis, Right Bicep Brachii Short Head, Right Anterior Deltoid, Left Anterior Deltoid, Right Triceps, Left Triceps, Right hand, Left hand, Right Brachioradialis]*, and the best eight muscles were selected based on specific criteria outlined in the previous section. The guitarist performed a set of standard technical gestures at a fixed tempo (100 BPM), and muscles with consistently low RMS values were excluded from further consideration. To refine the muscle selection process and identify muscle activation patterns associated with each guitar gesture, the ZCR was utilized to evaluate average activation levels within predefined time intervals. By combining the RMS and ZCR features, we determined the optimal set of eight upper limb muscles for accurately classifying each specific gesture. The **flexor carpi radialis**, **extensor carpi radialis**, and **brachioradialis**, were found to be important for wrist movement during guitar playing. The **right anterior deltoid**, responsible for shoulder movement, played a significant role in techniques such as strumming and chord changes. The **biceps**, involved in elbow movement, showed high activation during bending and vibrato techniques. The triceps and right brachioradialis were discarded, and the hand muscles exhibited low signal-to-noise ratio (SNR) due to cross-talk.

2.5. Proposed Model

2.5.1 Dataset creation

The dataset used in this study was created to address the lack of public surface electromyography (sEMG) datasets for guitar techniques. The dataset consists of sEMG signals captured during seven different guitar techniques, namely fingerpicking, strumming, bending, down picking, alternate picking, tapping, and pull-off/hammer-on. To ensure consistency across participants, each technique was paired with a corresponding guitar riff. Two acquisitions were performed for each technique, asking the tester to play at different levels of muscle contraction, in order to detect the maximum and minimum contraction to train the regression model. The dataset was collected from four guitarists with varying levels of expertise, each acquisition lasted for 30 seconds at a constant tempo of 100 beats per minute.
The signals were captured from eight selected muscles, and audio files were also saved alongside each acquisition. The dataset was labeled and organized using Raw Power software extracting the RMS feature from the raw signal. The processed acquisitions, along with a target dataset, that included class labels for each gesture, were loaded into a pandas data frame to shape the dataset for applying supervised learning. The dataset was split into an 80% training set and a 20% validation set for hyper-parameter tuning and model evaluation.

2.5.2 Model Desing

The model architecture designed for this study is based on a Bidirectional Long Short-Term Memory (BLSTM) layer followed by two fully connected Dense layers. The BLSTM layer, which incorporates a recurrent neural network (RNN) with LSTM cells, is known for its ability to capture temporal patterns in time series data. The bidirectionality of the BLSTM allows it to capture temporal dependencies in both the past and future directions, enhancing the model's predictive capabilities.
The model was implemented using TensorFlow (version 2.12.0) and the Keras API. To enhance the generalization and feature extraction capabilities of the RNN, two Dense layers were added after the BLSTM layer. The model also includes

a Dropout layer, which randomly drops out a certain proportion of neurons during training to prevent overfitting. The output layer utilizes the softmax activation function for multi-class classification, providing class probabilities for each gesture.

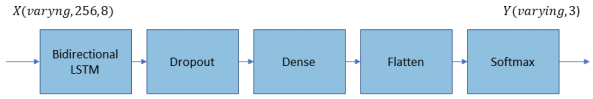


Figure 1: The model architecture diagram

The cost function used to optimize the model is the categorical cross-entropy loss, which penalizes incorrect predictions in multi-class classification problems. The weights of the model were updated using the ADAM optimization algorithm, a combination of gradient descent with momentum and RMS propagation. The learning rate, which controls the magnitude of weight updates during training, was selected to ensure optimal convergence and stability. Additionally, L1 kernel regularization was applied to the loss function to prevent overfitting by encouraging sparse weight values in the network.

3. Experimental Set Up and Evaluation

The development of this project was guided by different experiments; All of them were conducted using the same right-hand solid-body electric guitar, performed in a stationary standing position.
The protocol pipeline (as shown in Figure 3) for acquiring, processing, and classifying sEMG signals in real-time to control the guitar sound is summarized in these steps:

3.1. Protocol Pipeline

Prefiltering Stage: A prefiltering stage is applied to ensure signal fidelity. This stage consists of an anti-aliasing filter and a high-pass filter (with a cutoff frequency of 5Hz) applied in cascade.
Data Acquisition and Processing: The wearable board transmits the signals to a computer system running a proprietary data acquisition platform called Row Power. This platform performs crucial operations such as feature extraction and signal packaging. The packaged data is then simultaneously fed into two separate recurrent neural network (RNN) models.

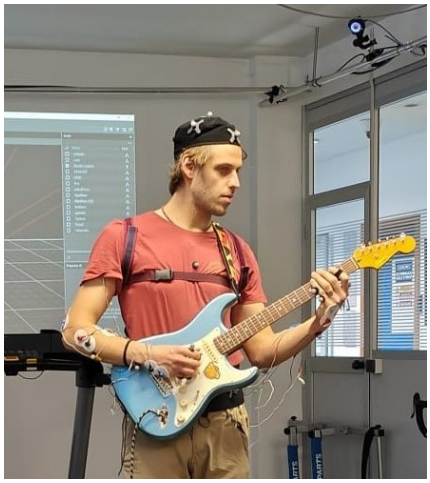


Figure 2: Guitarist during the acquisition session with electrode pads placed on arms.

Gesture Classification: The first RNN model is dedicated to gesture classification, facilitating the selection of pedalboard presets based on the recognized gestures.
Effect Modulation: The second RNN model operates as a regression model, enabling continuous modulation of the chosen effects based on the sEMG signals.
Max/Msp 8 Patch: The predicted parameters from the RNN models are sent to a Max/Msp 8 patch using the Open Sound Control (OSC) network protocol. The Max/Msp 8 patch encapsulates a series of five Virtual Sound Technology (VST) plugins arranged in a sequential manner to achieve the desired audio effects and modifications.
Audio Output and Processing: The audio output from the Max/Msp 8 patch is routed to a PA system. Ableton, a Digital Audio Workstation (DAW), is employed as an intermediary between the Max/Msp 8 patch and the PA system. Ableton provides a robust platform for audio mixing, signal processing, and playback control, ensuring high-quality sound reproduction.

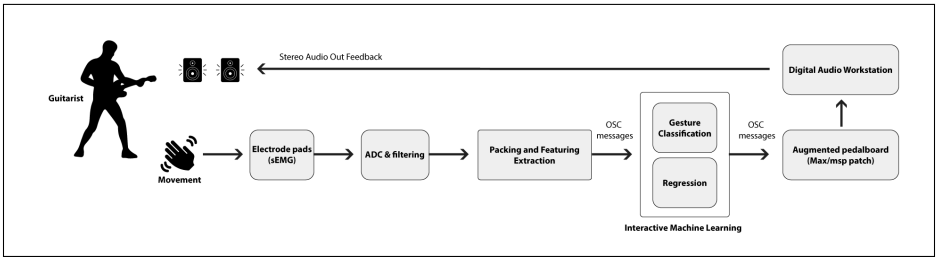


Figure 3: This image shows the entire signal flow, from its generation and acquisition from the user's body to the modification of the sound by the Max/Msp patch.

Layer	Output Shape	Number of parameters	Units
Input	(None, 256, 8)	-	-
BLSTM(kernel regularizer=l1(0.001))	(None, 256, 32)	3200	32
Dropout(0.1)	(None, 256, 32)	0	0
Dense (activation= 'tanh')	(None, 256, 64)	2112	64
Dense(activation= 'relu')	(None, 256, 32)	2080	32
Flatten()	(None, 8192)	0	0
Dense(activation= 'softmax')	(None, 3)	24579	3
Model Summary			
Total Parameters			31,971
Trainable Parameters			31,971
Output Shape			(None, 3)

Table 1: Architecture of the classifier model with 3 classes. Units indicate the number of neurons for each layer

3.2. Models training and evaluation

The training process involves two parts: one for the gesture classification and the other for the regression model. For classification, we define, in Max/Msp, four sonic presets associating them with an intention among calm, happy, and frenetic; we associate each class to the corresponding guitar gestures (among the seven guitar techniques defined in section 2.5.1) used to convey a specific intention. The model classifies the gestures and triggers the preset change in real-time. In the second part, we train the regression model; it modulates a group of guitar effects based on the amount of muscular contraction. The regression model has one output for each parameter, changing them via OSC messages during the performance. It allowed the guitarist to modulate filter parameters by adjusting muscular effort. The model learned the relationship between effort and parameter changes through multiple acquisitions of gestures performed with different levels of contraction.

3.2.1 Models evaluation

We present the final architecture of the classifier model in Table 1. The architecture of the regression model is identical besides the output shape of (None, 6) having 6 pedalboard parameters to control, and the output activation function equal to ReLu.

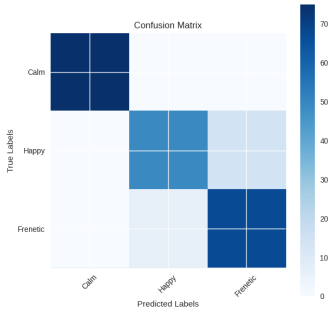
In Table 2 we present the chosen hyperparameters, after a manual search to strike a balance between model complexity and generalization capability, taking the model's parameters low for the real-time application. We present the training validation plots in Figure 5, 4 for the two proposed models; during training, we monitored the loss function and accuracy to ensure convergence (Fig. 5). Finally, we evaluate the model on the test set by plotting the confusion matrix and regression residuals (Fig 4).

3.3. Final Experiment

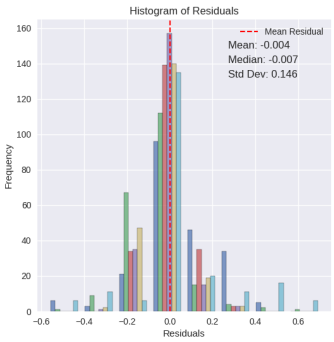
For the final evaluation, we had an experienced guitarist try the system. We evaluated the system in a performance scenario, after the training

Hyperparameter	Value
Input Time Steps	256
Regularization Strategy	L1(factor= 0.001)
Dropout Rate	0.1
Optimization Algorithm	Adam
Learning Rate	0.0001
Loss Function	Categorical Cross Entropy
Batch Size	32
Early Stopping	Monitor 'loss function' with patience 10
Early Stopping	Monitor 'validation loss' with patience 10

Table 2: Model's Hyperparameter Values

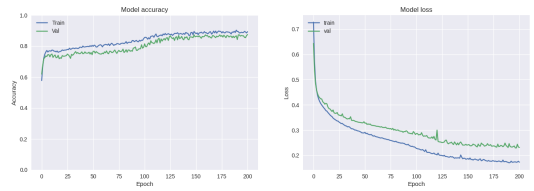


(a) classifier's confusion matrix

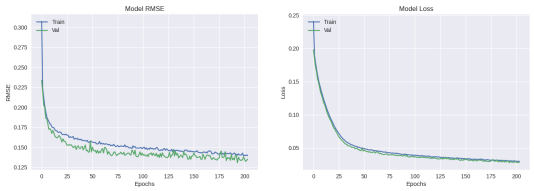


(b) regression residuals

Figure 4: (a) Confusion matrix of the classifier with three classes over the test dataset: overall accuracy 0.905. (b) Histogram of the regression's residuals, the bins' Gaussian distribution shows the effectiveness of the model.



(a) classifier's loss and accuracy



(b) regression's loss and RMSE

Figure 5: (a) Evaluation plot of the classifier's training: loss 0.1801 - accuracy 0.9245.(b) Evaluation plot of the regression RNN, with: MSE loss 0.0316, RMSE 0.1464

process to adapt the two RNNs according to the subject. To collect the musician's perspective, we design a questionnaire following the typical DMI evaluation methodologies. The questionnaire results are publicly available in the full text of the paper which this summary refers to.

4. Conclusions

The objective of this paper is to propose an innovative interaction method based on the conjunction of electromyographic signals and deep learning analysis, integrated into a DMI for guitarists to develop a tool that musicians may use to sonify their gestures during the performance. With the proposed data acquisition protocol we were able to select the best muscle groups for guitarist gestures classification, the results were used for the implementation of the proposed tool, which is able to modulate a set of effects accordingly to the user's muscle activation signals, consequently changing the sound of the guitar during a performance. We have described how to integrate the sEMG signal into the development of a DMI for guitarists, but applicable to many other instruments, by using a custom Gestural Interface able to handle full body mapping. Finally, an evaluation strategy based on a questionnaire has been presented, with the goal of paving the way for other researchers interested in integrating muscle signals into artistic performances.

References

[1] M. Barbero, R. Merletti, and A. Rainoldi. *Atlas of Muscle Innervation Zones: Understanding Surface Electromyography and Its Applications*. Springer Milan, 2012.

[2] Alejandro Toro-Ossaba, Juan Jaramillo-Tigreros, Juan C Tejada, Alejandro Peña, Alexandro López-González, and Rui Alexandre Castanho. Lstm recurrent neural network for hand gesture recognition using emg signals. *Applied Sciences*, 12(19):9700, 2022.

[3] Federico Ghelli Visi and Atau Tanaka. Interactive machine learning of musical gesture. *Handbook of artificial intelligence for music: Foundations, advanced approaches, and developments for creativity*, pages 771–798, 2021.