Machine learning for recognition of individuals from motion capture time series: performance and explainability

Elena Mariolina Galdi, Dipartimento di Ingegneria, Università di Ferrara
Marco Alberti, Dipartimento di Matematica e Informatica, Università di Ferrara
Alessandro D'Ausilio, Dipartimento di Neuroscienze e Riabilitazione
Alice Tomassini, Istituto Italiano di Tecnologia, CTNSC@Unife

Outline

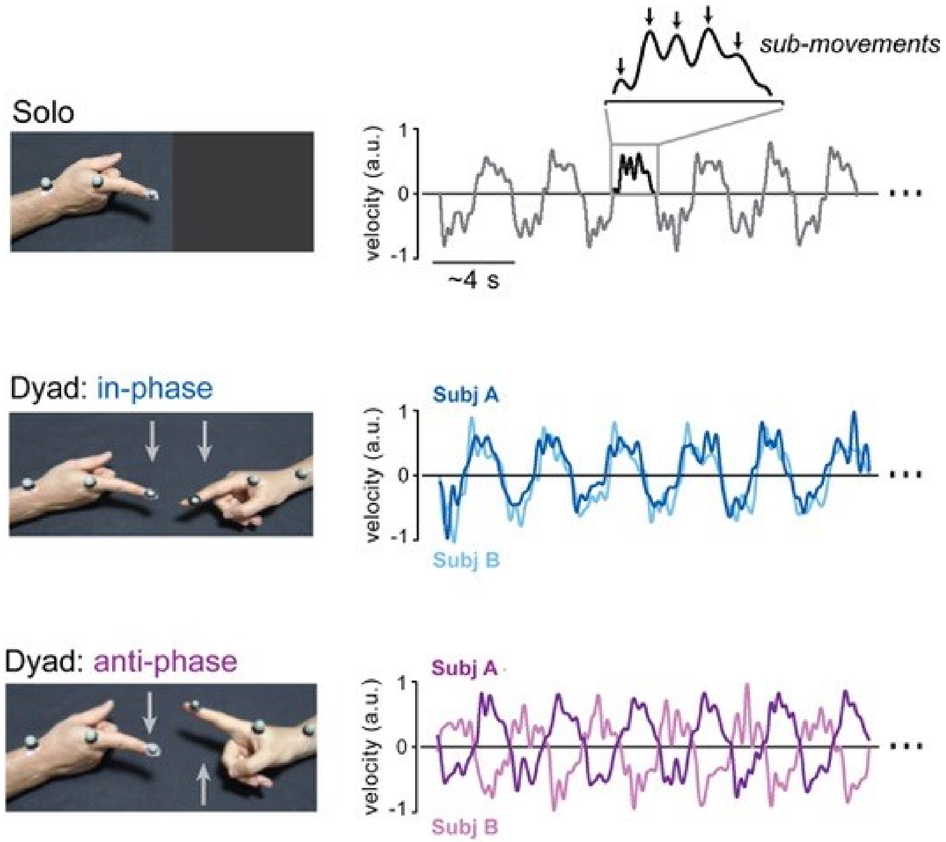
- Motion capture dataset: time series of 60 people performing simple exercises with their index fingers
- A Convolutional Neural Network can recongnize the person with 75% accuracy
- Why?
  - Impact of preprocessing
  - Do computational findings agree with the neurophysiological evidence?
- Ongoing work

Individual Motor Signature

Explainable AI

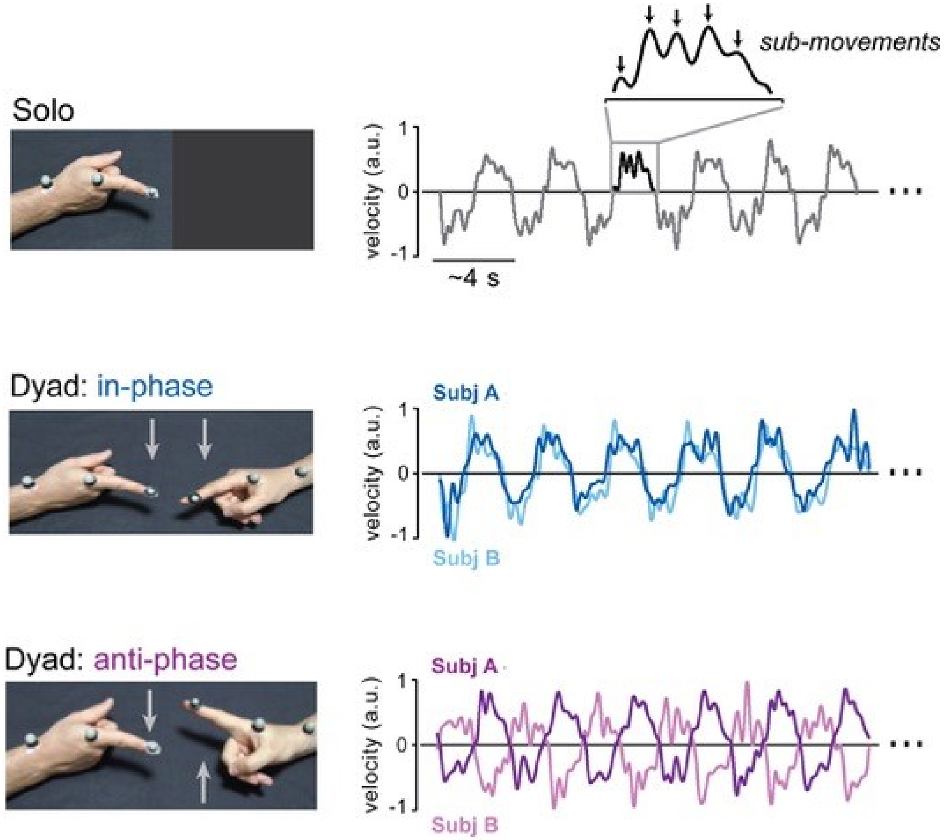
DATASET

Tomassini et al. , iScience, 2022
Interpersonal synchronization of movement intermittency



- 60 partecipants forming 30 couples
- Reference metronome set @ 0.25 Hz
- Sampling rate = 300 Hz
- Trial's duration = 2,5 minutes
- Total points for each time series = 45000
- Only X coordinate of fingertip movement recorded

Tomassini et al. , iScience, 2022
Interpersonal synchronization of movement intermittency



- Submovements: tiny corrective bumps in the speed profile in the range of 2-3 Hz
- Previous research points out that submovements have a crucial role in the interpersonal synchronization
- ***Can these microscopic movement characteristics be used for the identification of individual movement fingerprints?***

CAN AN INDIVIDUAL BE RECOGNIZED FROM HIS/HER FINGER MOVEMENTS?

Convolutional Neural Network (CNN)

- To investigate whether it is possible to identify a subject from the index finger extension and flexion, we used a CNN.
- The choice of a CNN for multiclass classification, including of time-series, has been shown to be effective*.
- Optimizer : RMSprop
- Regularization techniques: Early Stopping and Dropout 0.25

Convolution 1D	•Kernel: 3@64 •Act Funct: ReLU
Convolution 1D	•Kernel: 3@64 •Act Funct: ReLU
Max Pooling 1D	•Pool size:2
Convolution 1D	•Kernel: 3@128 •Act Funct: ReLU
Convolution 1D	•Kernel: 3@128 •Act Funct: ReLU
Max Pooling 1D	•Pool size:2
Dense	
Dense	
Softmax	

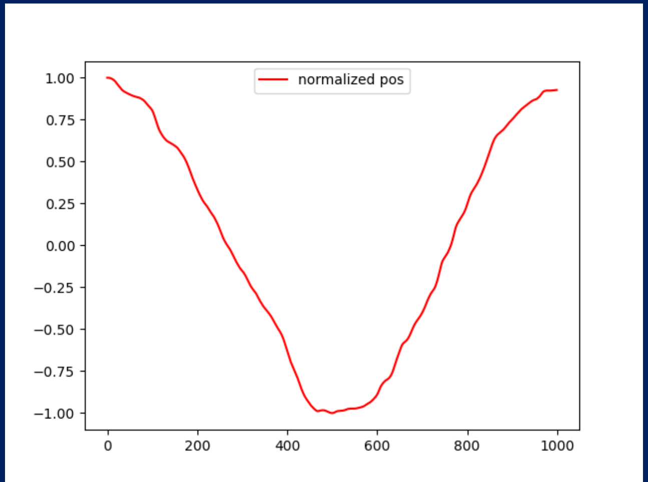
* Cui, Z., Chen, W., Chen, Y.: Multi-Scale Convolutional Neural Networks for Time Series Classification (May 2016)

Preprocessing

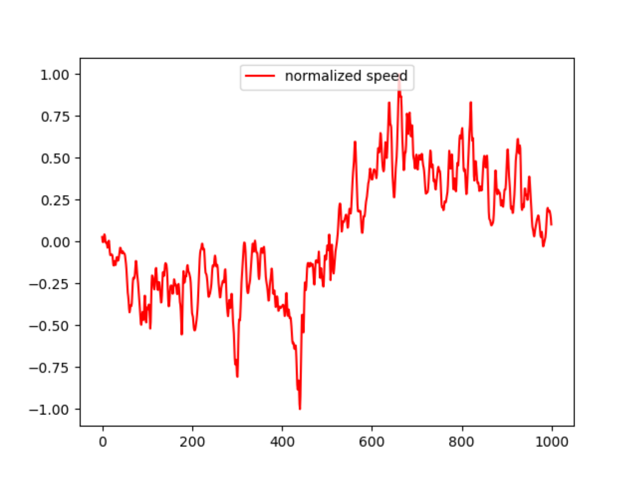
Series Type	• Position • Speed • Acceleration
Filtering Method	• MAW • Band Pass
Cutting	• Extension-Flexion • Sliding Window
Normalization	• ON • OFF

Series type

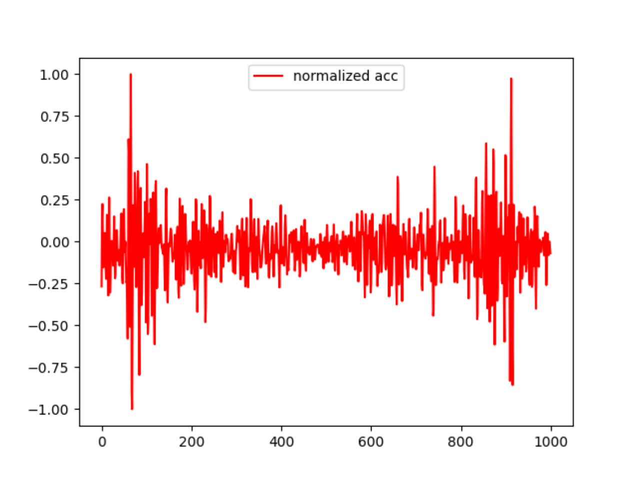
Position



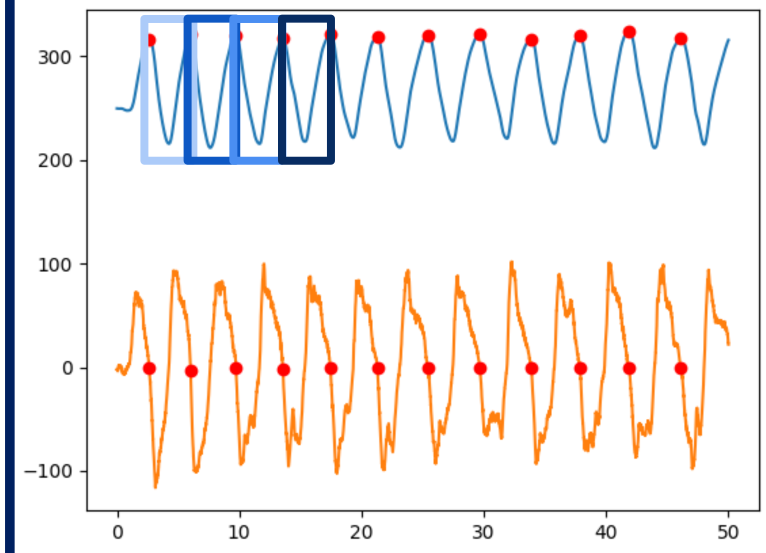
Speed

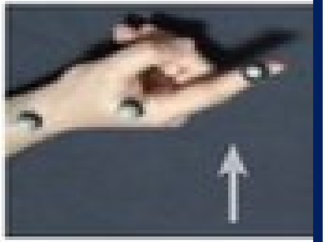


Acceleration



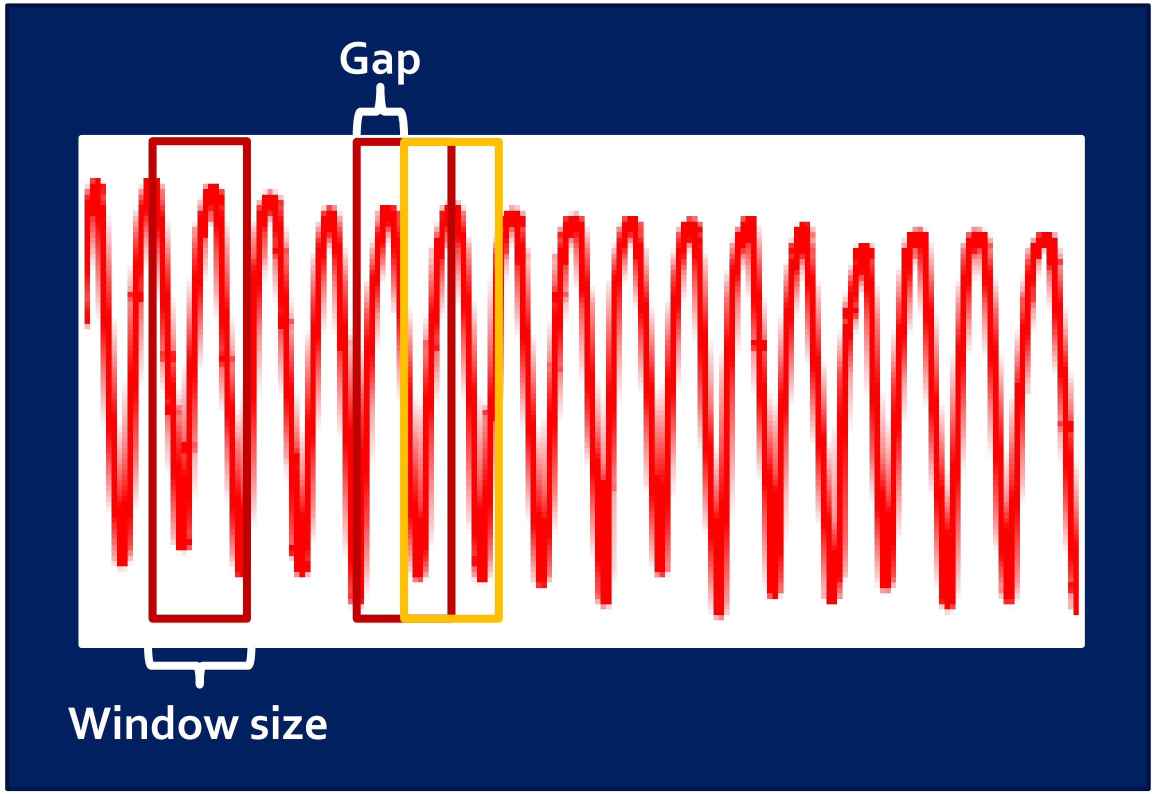
Cutting 1/2





- **Extension & Flexion**:
  - Cut the time series corresponding to the maximum finger positions on the x-axis
  - Each sub-series represents the complete movement, extension, and flexion of the index finger
  - Resizing is necessary

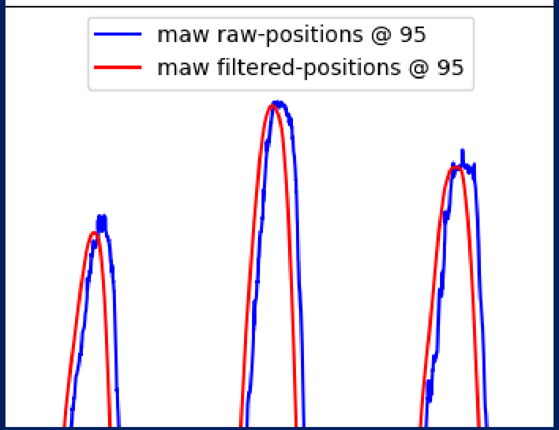
Cutting 2/2



- **Sliding Windows** :
  - Fixed windows size
  - Fixed gap between two consecutive subseries

Windows size = 5

Windows size = 95

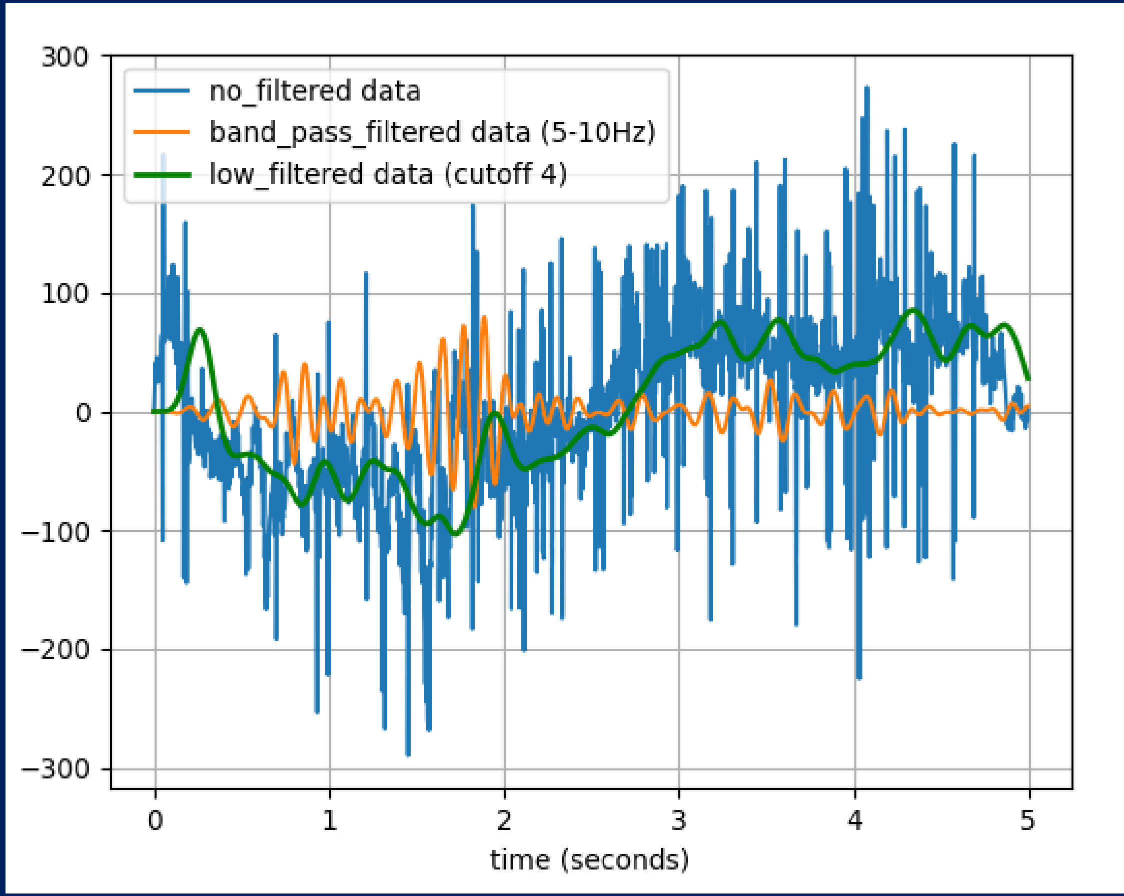


Filtering Method 1/2

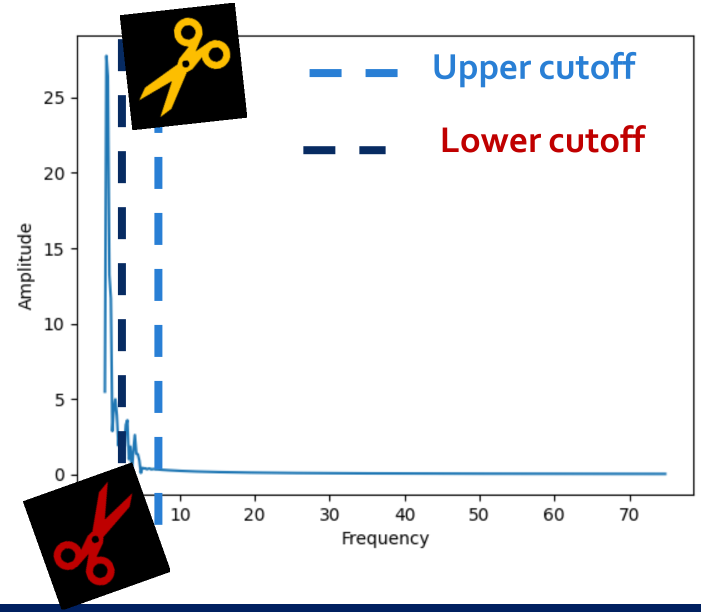
- **Moving Average Window** :
  - New series where the values are comprised of the average of raw observations in the original time series.
  - **Windows** size defines the number of raw observations used to calculate the **Average**
  - **Moving** : the window is slid along the time series to calculate the average values along the series
- A large window size drives away from the raw value, but it maintains the main signal component, the same trend.

Università degli Studi di Ferrara

Filtering Method 2/2



- **Band Pass Filter** :
  - Filter in frequencies domain
  - Set the upper and lower cut-off frequency points



Results

Preprocessing setting:

Series type: speed

Filtering method: MAW (window's size = 5 pts)

Cutting choice : one complete finger movement

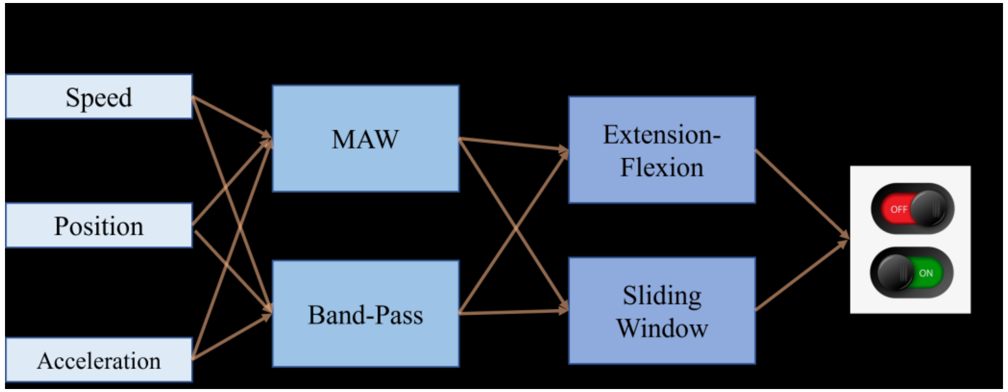
Normalization : ON

Accuracy : 75 %

Number of subjects: 60
The baseline accuracy of a random classifier is **:**
1/60 = 1,7%

WHICH ARE THE MOST RELEVANT CHARACTERISTICS FOR IDENTIFICATION?

Modular Structure



Idea: examine how the accuracy of our CNN is affected by the choices made in the data preprocessing

We now present the most significant results we obtained from this approach.

Series type

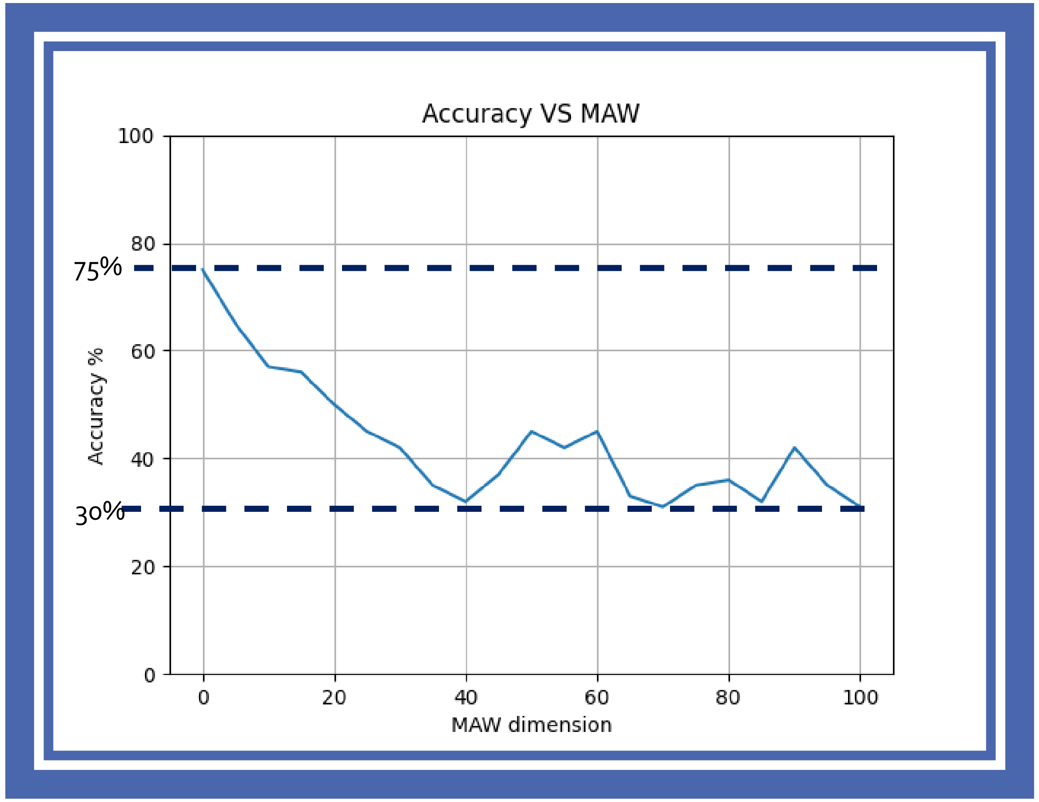
Series type	Accuracy
Position	35 %
Speed	75 %
Acceleration	65 %

The accuracy for the different series types was calculated using a low-pass filter set at 50 Hz and cut based on the maximum finger position

Cutting choice

- Extension&Flexion   VS Sliding Window:
  - More data from sliding windows
  - But close in terms of resulting accuracy

Filtering Method: MAW

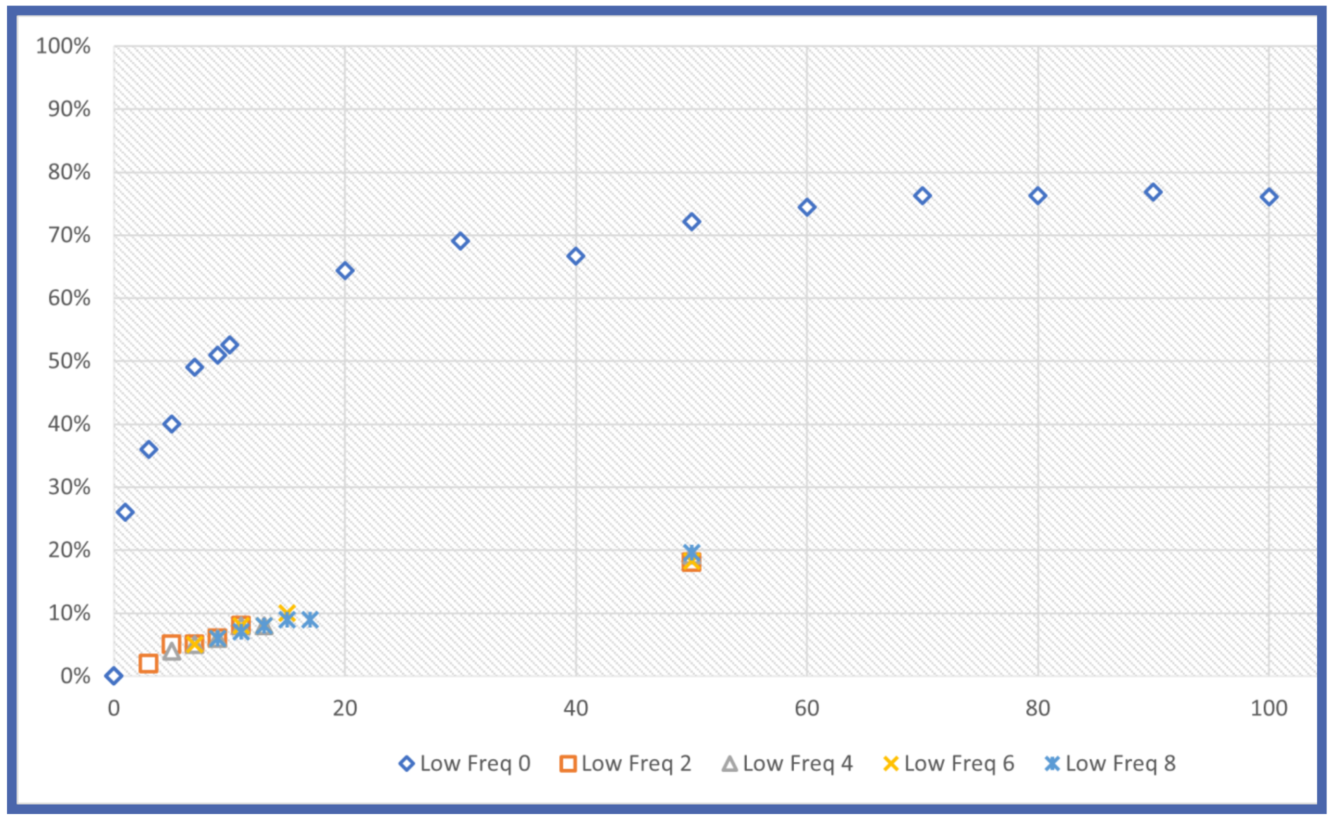


- Remarks :
  - Max Accuracy = 75%

    ⇒ One-sized Window: no filtering applied
  - Min Accuracy = 30%

    ⇒ Large window: only the main signal component

Filtering Method: Band Pass



- The lower cutoff is indicated in the caption, while the upper cutoff is shown on the x-axis.
- Remarks:
  - The fundamental frequency (0.25 Hz) is the most meaningful frequency
  - From to 20 Hz onwards, there is no physiological relevance anymore
  - @20 Hz : accuracy = 65%

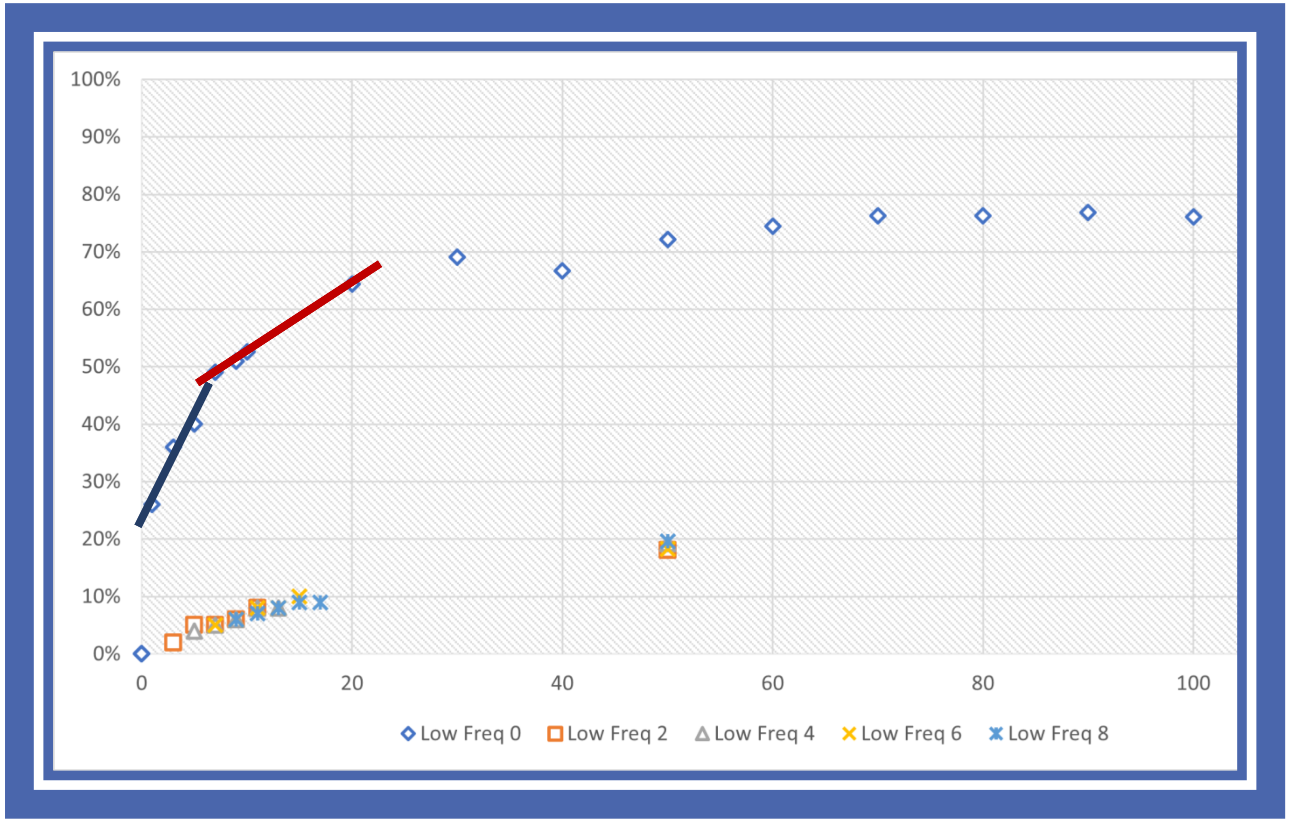
    ⇒Still remarkable result when compared with 1,7% random guess

CONCLUSIONS

Main points

- It is possible to recognize subjects by their index finger movements with up to 75% accuracy vs. a baseline accuracy (random classifier) of 1.7%
- The fundamental harmonic is the pivotal aspect in the recognition of subjects, but not alone
- Higher frequencies contribute significantly to an increase in accuracy, but only in the presence of the fundamental frequency.

Submovements

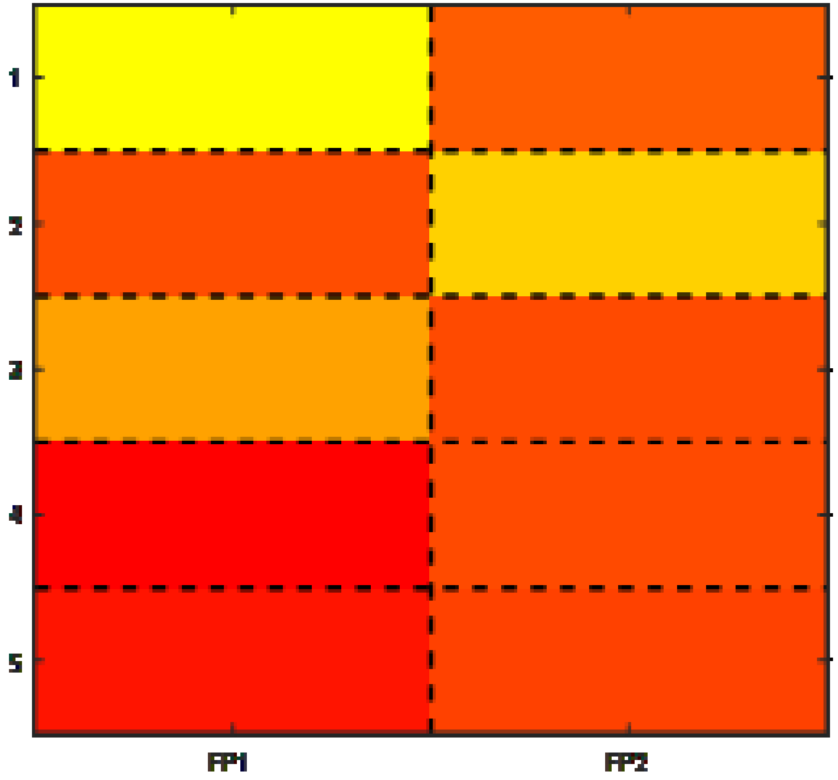


- Initial goals was to investigate the role of sub-movements (2-3Hz):
  - No relevant contribution when considered alone
  - If added to the main harmonic, they cause significant improvement (more so than the one obtained adding frequencies over 10 Hz)

ONGOING AND FUTURE WORK

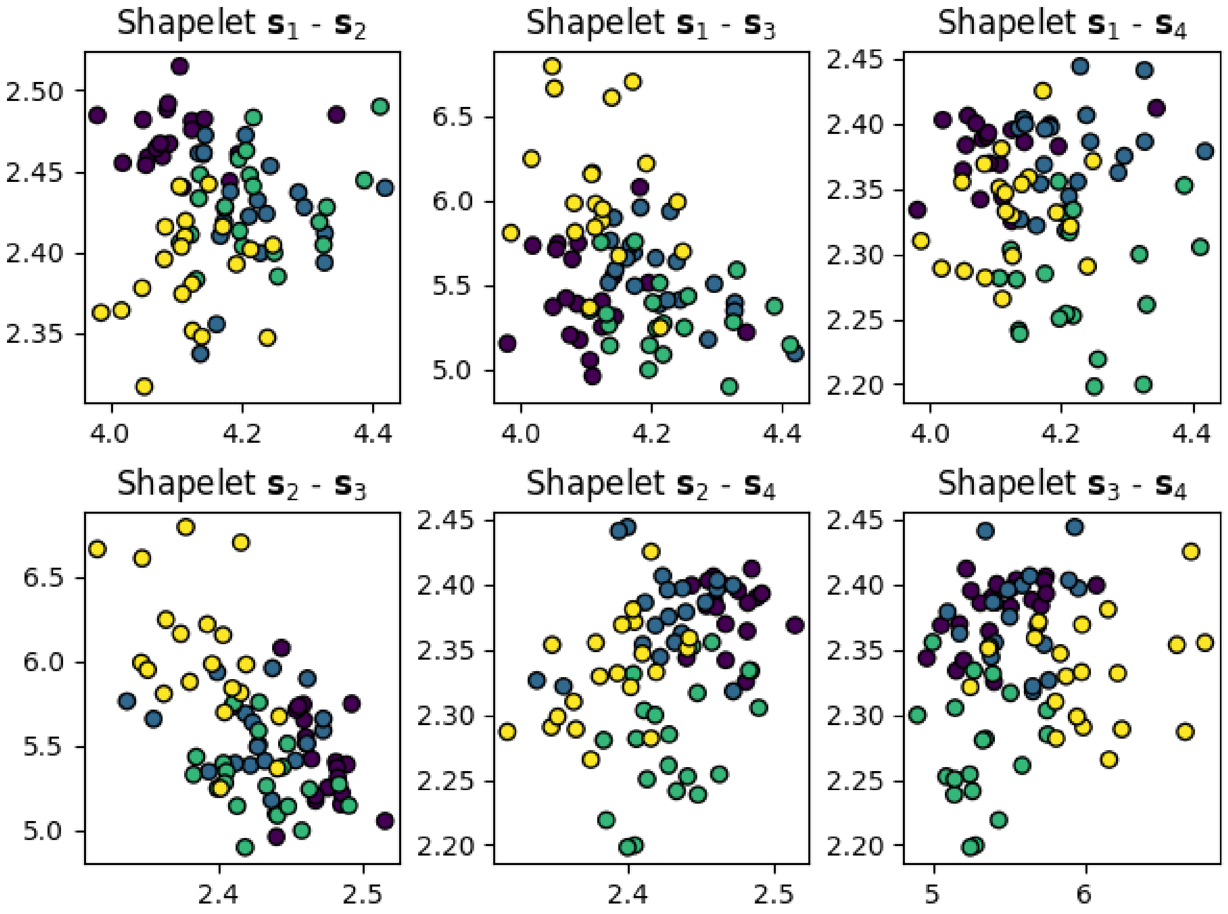
Parsimonious Linear Fingerprinting (PLiF)

1. Fit a Linear Dynamical System (LDS) on the collection of m sequences, and
2. extract a few meaningful features ("fingerprints") out of the LDS
3. map each time series on the space of the fingerprints, for e.g. classification or clustering



Shapelet learning

- sub-sequences of values that are most representative of class membership
- subsequences that maximize the information gain when dividing the set of all subsequences into two classes based on their distance from the candidate



- 4 classes, each color represents a class
- learned 4 shapelets
- each diagram shows time series in the shapelet-space indicated in the legenda

Future work

- Investigate time series captured from the same subjects performing different exercises (individual motion signature)
- Investigate the effect of paired vs. solo performance (do time series by two subjects become closer when they act together instead of separately? also wrt. submovements)