University of Padova

School of Engineering

Master Degree in Telecommunication Engineering

Analysis techniques for expressive movement

Comparing conductor movements with listeners gestures. A case of music-mediated imitation?

Candidate

Chiara Bizzotto

Supervisors

Prof. Giovanni De Poli

Prof. Marc Leman

Assistant supervisor

Dr. Pieter-Jan Maes

Academic Year 2014/2015

Contents

Abstract

In the proposed experiment, we want to investigate whether correlations exist between the conductor gestures while conducting an orchestra, and the movement responses of people, merely listening to the musical outcome of the orchestra performance. Additionally, we would like to assess whether people familiarity with the music contributes to this correlation. The main purpose of my thesis is to investigate several analysis techniques that one can use to treat expressive gestures signal.

Introduction

The main task of a music conductor is to temporally coordinate a musical ensemble performance. Therefore, a conductor typically uses expressive gestures to inform musicians about his/her musical goals and interpretation. From the side of the listener, research has demonstrated that listening to music induces body movements that convey how people interpret and perceive musical expressiveness.

The pillar behind this research is the existence of a *model of musical communication* (Leman, [1]) in which the transmission of intention is possible through the encoding and decoding of bio-mechanical energy (the playing of an instrument).

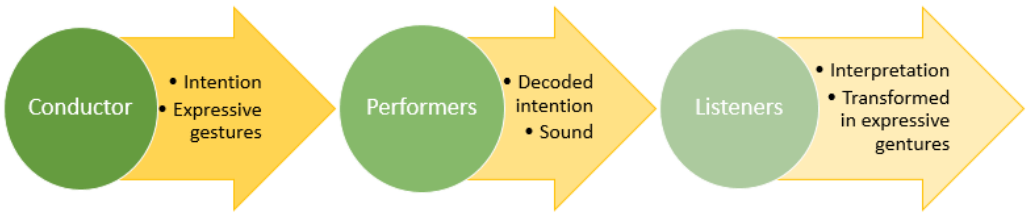


Figure 1: Scheme of the model of musical communication

- It all starts with the conductor: he/she gives the orchestra instructions (*expressive gestures*) to deliver his/her interpretation of the score;
- the orchestra has then to translate these instructions into a musical outcome (*sound*) using both the human body and a mediation technology (musical instrument);
- the instrument transforms part of the bio-mechanical energy into sound energy and part into haptic energy, that returns through the touch to the performer (see Fig. 2);

- for this experiment only the sonic energy is then delivered to the listener (in fact the participants can't see the conductor or the orchestra playing);
- through mirror processes, the listener can understand the music intention although it can be different from the conductor/performers one;
- the understanding is then processed and expressed through movements (*expressive gestures*).

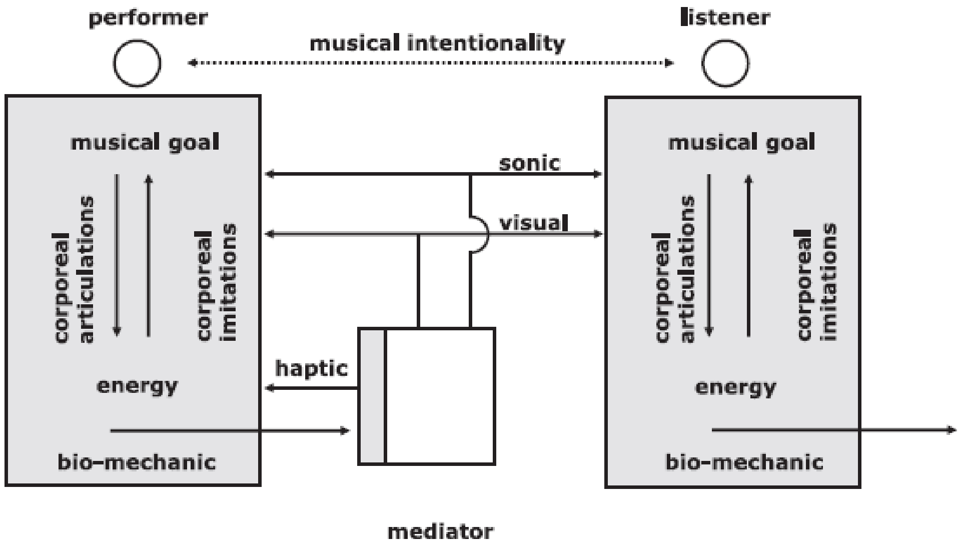


Figure 2: Detail of the general model between performer and listener. Figure from [1].

The aim is first to see if there is a connection between the expressive gestures of the conductor and the ones of the listeners, as they should be an encoding of the same intention. Then I want to check if familiarity with the music will modify this correlation. During the experiment there will be in fact a phase where subjects will learn the music listening to it multiple times. We know in fact that the repetition of an experience creates a link between it and a peculiar motor action that leads to a development of an internal model (see [2]). The hypothesis is in fact that, due to familiarity, expression-responding gestures will change:

- for excerpt with higher familiarity they will become intention-driven (listeners movements will come closer to conductors movements), and have a higher group-commonality;

- for the others they will be less intention-driven, but higher than in pre-test.

I will dedicate the first Chapter of my thesis to briefly introduce the reader to the experiment we performed: the equipment, the participants and the music stimuli (Chapter 1).

I have then divided the rest into two main parts. Part I is dedicated to the pre-processing that is a sort of data manipulation that is required before starting the analysis: in particular I will treat the inconvenience of the "False-values" in Chapter 2 and the necessity of a preliminary smoothing in Chapter 3. Finally in Chapter 4 I will explain how to obtain the useful signals of speed and acceleration from the raw data.

Part II instead, will treat the main argument of my thesis that is the analysis. First, in Chapter 5, I will talk about the Functional Principal Component analysis, both on a theoretic level and on a more practical one: I will do a brief description of the principal components of the data and then I will use the results to compare participants and conductors.

In Chapter 6 I will look directly at the plots of the data and I will deduce interesting conclusions on the influence of familiarity both on the amplitude of the movement and on the concept of anticipation.

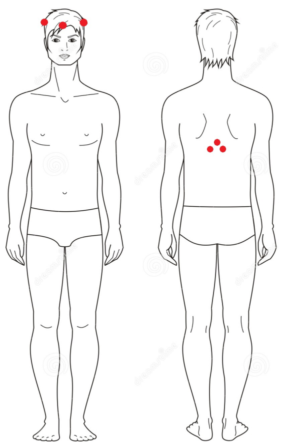
I will talk about the correlation analysis in Chapter 7 where I will treat both the intra-group correlation and the one with the conductor.

Finally in Chapter 8, I will introduce briefly the possibility of using Mutual Information to check for correspondences between the signals.

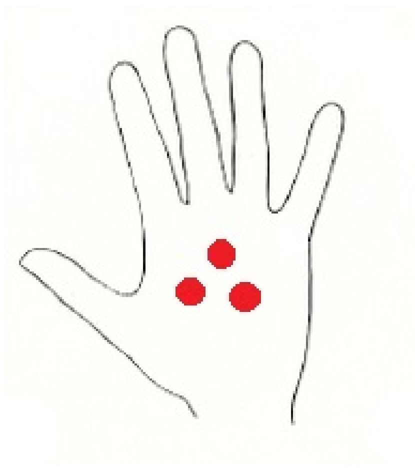
At the very end of my thesis I will try to deduce some final conclusions of my work, underlying some problems and some suggestions for a further analysis.

1. Procedure and stimuli

The experiment is set in a circular environment where a Motion Capture system is in action at sampling frequency of $100Hz$. Each participant is seated on a high chair, that recalls the position of the actual conductor, for the whole duration of the tasks and he/she has to wear some markers: on both hands, on the back and on the head (see Fig. 1.1).



(a)



(b)

Figure 1.1: Position of the markers on the head and on the back (a) and on both hands (b).

The stimuli chosen are three excerpts from the opera "*Don Pasquale*" of Gaetano Donizetti performed by two different conductors (Fig. 1.2): Riccardo Muti (M1, M2 and M3) and Carla Del Frate (DF1, DF2 and DF3).

The peculiar excerpts are chosen for practical reasons in addition to analysis ones:

- firstly it is necessary to pick segments for which the conductor movements (acquired through motion capture system in a previous time) don't have missing values;
- lastly there have to be some important differences between the two executions of the same excerpt and possibly some differences in style in the excerpt itself.

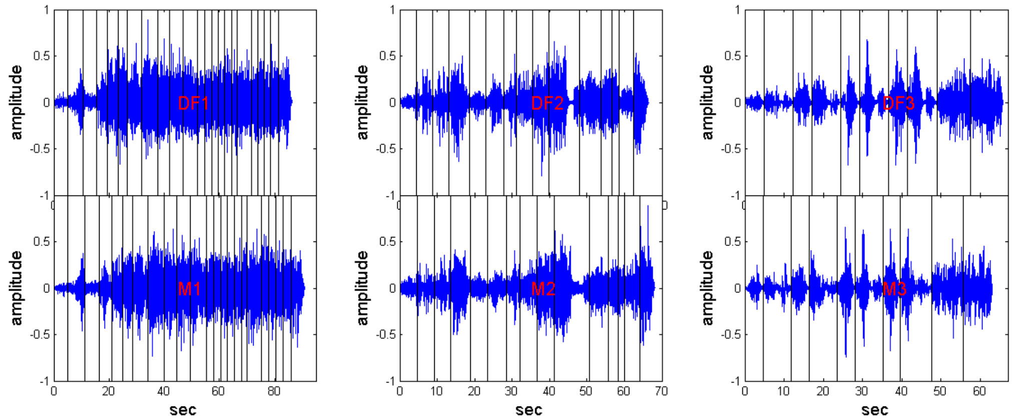


Figure 1.2: Amplitude plots of the three excerpts executed by the two conductors. The beginning of the bars are highlighted.

As we can see in Figure 1.2, there are some differences between the two executions especially in timing: for excerpts 1 and 2 the Muti signal has some delay in respect to the Del Frate one, while for excerpt 3 it happens the contrary. The black vertical lines represent the beginning of the measures and one can see that, while they start together (Muti and Del Frate), towards the end they are slightly out-of-sync. Furthermore one can see that, while excerpt 1 has quite a fast tempo (the distance between the bars is really small), excerpt 2 and 3 are slower and, in particular, one can notice that in excerpt 2 there is an increase of the tempo towards the end as the bars get more crammed.

For what concerns the magnitude of the amplitude, the differences are less visible. As I have highlighted in Figure 1.3, for excerpts 1 and 2 there is a moment in time when the Del Frate segment is quite evidently over the Muti one; however for excerpt 3 one can notice some more interesting discrepancies. In the central part of the signal it is evident that there are three repetitions of the same melody with different intensities: while for the

Del Frate excerpt(blue curve) there is a *crescendo* in the amplitude, for the Muti excerpt (red curve) there is an increase between the first and second repetition but then the intensity remains the same.

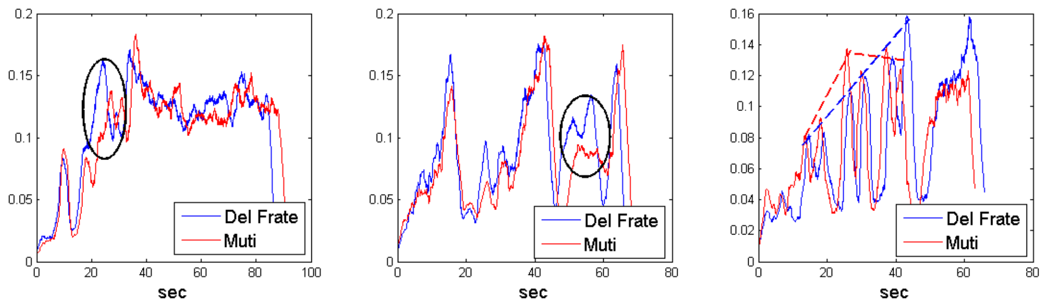


Figure 1.3: Envelope of the magnitude of the music excerpts: comparison between Muti and Del Frate

The group of participants is formed by 32 people, half female and half male, of age between 18 and 35 years and with some formal musical background, because they have a stronger auditory-motor couplings. Before the beginning of the experiment they are sorted into two gender-balanced groups of 16 people and, to each group, one of the two conductors was assigned: therefore we have the "Muti Group" and the "Del Frate Group". The difference between the two groups lays on the different training they get as I will now explain.
In fact the experiment consists mainly of three parts (Fig. 1.4):

1. pre-test;
2. training phase;
3. post-test.

For each part the participants have to do the same task:

> *To move along the music with hands and arms acting like they were the conductor, not using technical gestures or instructions but expressing the general feeling.*

Pre-test

In this first part each participant has to move along all six excerpts one time. The movements are recorded using the optical motion-capture system. The lights are off so that the participants can feel more comfortable and less embarrassed in their movements.

Training phase

The aim of the training phase is for the participants to become familiar with the three excerpts corresponding to the one particular conductor their group is assigned. Participants are informed about this goal, but they don't know about the two different conductors. For that purpose, we instruct participants to listen and move in response to the excerpts. This is similar to the previous phase, although only excerpts of one conductor are used and the movements are not recorded. Each excerpt is repeated four times in total: this number has been chosen because the subjects have to listen to the excerpt a sufficient number of times to learn it, without being so tired for the post-test that it influences their movements. Afterwards, all excerpts are played twice more, and participants are asked to merely listen to the music, without performing movements so that they can be more concentrated on the music itself than on the gestures. For this part the lights are on but dim.

Post-test

In this third part participants are asked again to move along the music as in the pre-test: all six excerpts from both the conductors are played once. The lights are off again.

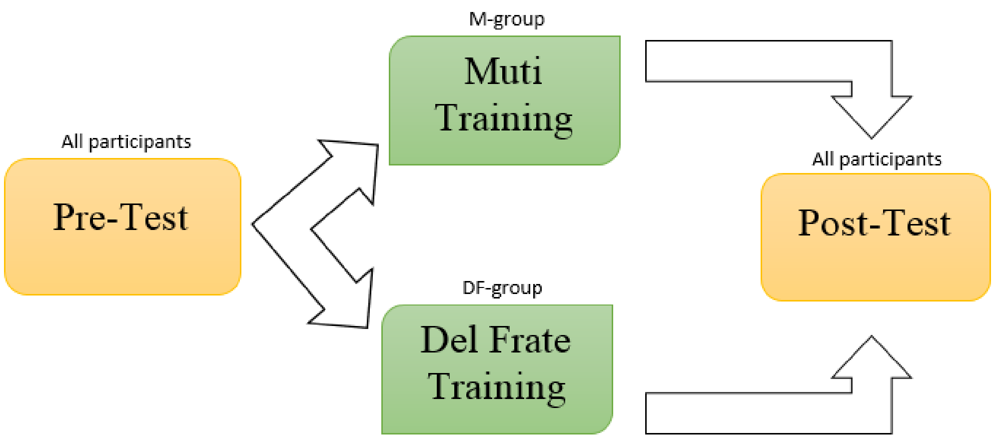


Figure 1.4: Scheme of the experiment

Part I

Pre-processing

2. “False-values” and interpolation

Once the experiment is completed for all the participants I have to deal with a huge amount of data: in fact the Motion Capture system produces a positional signal that is a 3D position *versus* time signal. Therefore for every person (32 in total) there is one 3D-signal for each marker (12 in total) for each excerpt (12 in total).

The system is set up in such a way that a *sync* signal is recorded as soon as the music starts so that I can be sure that all signals of all participants are synchronized. From the *sync* signal (Fig. 2.1) I have extracted the exact beginning of the music in term of samples and I can take out the part of the movement signal corresponding to the musical excerpt.

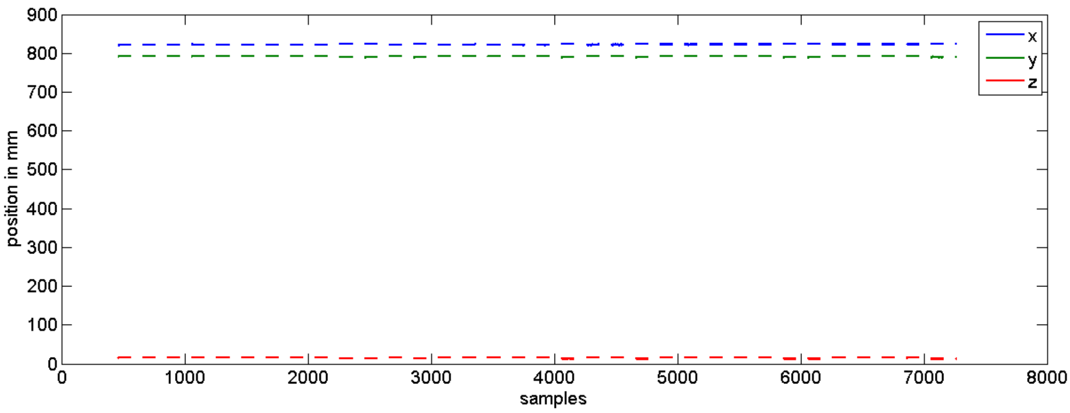


Figure 2.1: Example of sync signal

Once I have the movement signal of the same length as the musical excerpt and synchronized, a new problem is presented: in fact, during the acquisition of the data, some technical problems seem to have occurred and the signals present some “jumps” as one can see in Figure 2.2.

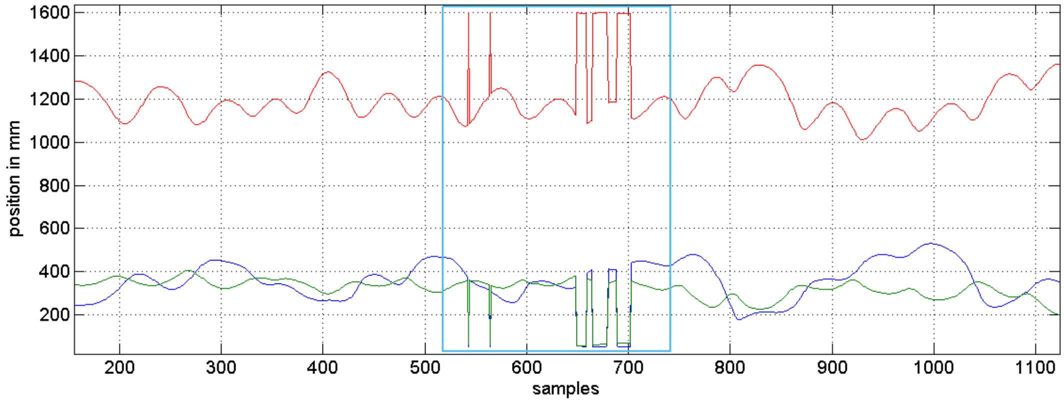


Figure 2.2: Example of "false value".

These values deviate too much from the signal and appear as discontinuities: because I am dealing with human movement and because the signal represents position (hence it should be continuous), I can consider them errors. I have decided to treat them as missing values (`NaN`) and then interpolate over these gaps to obtain a whole signal.

Because they seem to appear randomly in the signal, to find them I proceed as follows (see Listing 2.1):

- I fix a threshold as a percentage of the standard deviation of the signal, depending on the marker (the hands have more variation in respect to the head or the back).
- I set the first non-`NaN` value of the signal as *reference value*.
- I compare it with the next value:
  - if their difference is under the threshold, the sample is okay and it is set as the new *reference value*,
  - otherwise it is marked as "false value" and substituted with `NaN`: the *reference value* isn't changed.

```
for k=1:3
        reference(k) = data(find(~isnan(data(:,k)),1,'first'),k);
        %initialization of reference value = first
        %non-NaN value of signal
        for i=1:length(data)
            if abs( data(i,k)- reference(k) ) <= par*dev_st(k)
                %if the difference with the ref
                %value is under the threshold new
                %ref value is set and the sample
```

```
            %is okay: no false-value
            reference(k) = data(i,k);
            temp(i,k) = data(i,k);
        else
            %otherwise the ref value does not
            %change and a NaN is put instead
            %of the false-value
            temp(i,k) = NaN;

        end
    end
```

Listing 2.1: Code for finding the “false values”

After I found all the “false values” and I put `NaN`s in their place, I need to interpolate over the new missing values because my analysis techniques don’t behave well with `NaN`s.

Because interpolation doesn’t work if `NaN`s are at the beginning or at the end of the signal, I decide to run the algorithm only from the first until the last non-`NaN` value. By doing this I don’t loose any information because anyway I should have discarded the beginning and the end of the signal. In fact in those parts the movements of the participants are not exactly expressive as people are just beginning to move or they sense that the music is ending and they stop moving earlier. I will take this aspect into account when I do the analysis.

I choose to use a cubic `spline` interpolation to gain in precision without losing too much in complexity: for precaution I set the limit of the gap that can be interpolated at 70 samples (0,7 seconds) because if I attempt to reconstruct a bigger interval I would be facing some lack of precision in respect to the original movement.

For the analysis I need only one signal for each body part, so I calculate the barycenter between the three markers as the mean of each dimension.

$$position(t) = \left(\frac{s_{1,x}(t)+s_{2,x}(t)+s_{3,x}(t)}{3}, \frac{s_{1,y}(t)+s_{2,y}(t)+s_{3,y}(t)}{3}, \frac{s_{1,z}(t)+s_{2,z}(t)+s_{3,z}(t)}{3} \right)$$

3. Smoothing

Because all the experiments were executed in a long span of time, I can't be sure that the noise conditions are the same for all participants and for all excerpts. Inspired by the work of Desmet *et al.* in [3], I proceed to do a preparatory smoothing on the positional data.

My idea, not having at disposal the noise of the system, is to bring all the signals to the same noise level. Therefore I proceed as follows, considering only one dimension:

- First I use a Moving average filter on the signals cycling on the size of the window (considering only odd numbers as I use a central window).
- Then I subtract the smoothed signal from the original one, obtaining a sort of noise.

$$n(win) = sig - sig_{smooth(win)}$$

- I model it as a Gaussian random variable an then I extract the standard deviation.

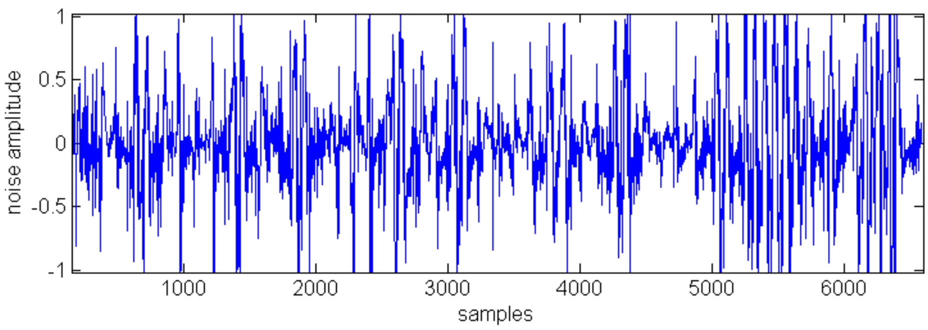


Figure 3.1: Example of noise with window size of 5 samples

- Lastly I plot for each excerpt the standard deviation of the noise varying in respect to the window of the smoothing filter for all participants (see Fig. 3.2).

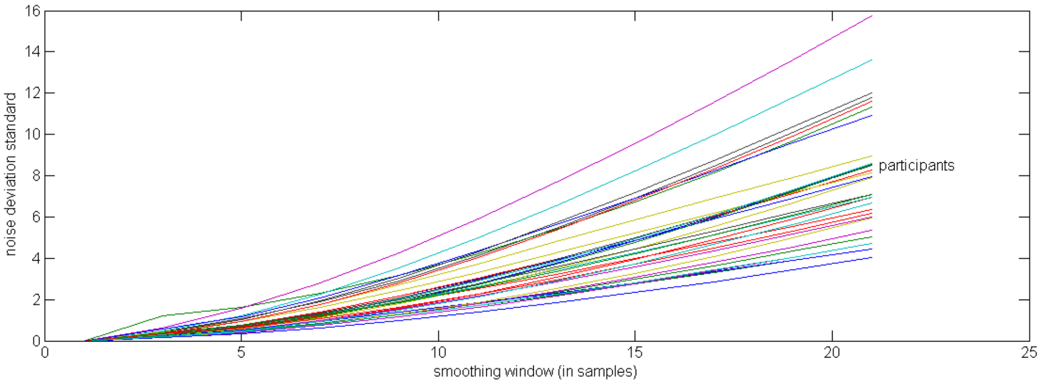


Figure 3.2: Standard deviation vs window size curves

For each participant I choose the window size that produces a standard deviation of the noise equal to 3, that is the lowest possible applicable to all participants and all excerpts (Fig. 3.3). Once decided the windows sizes, I apply the corresponding Moving Average filters to the positional data. I choose a version of the filter that behaves well with `NaN`s and that preserves the extremes.

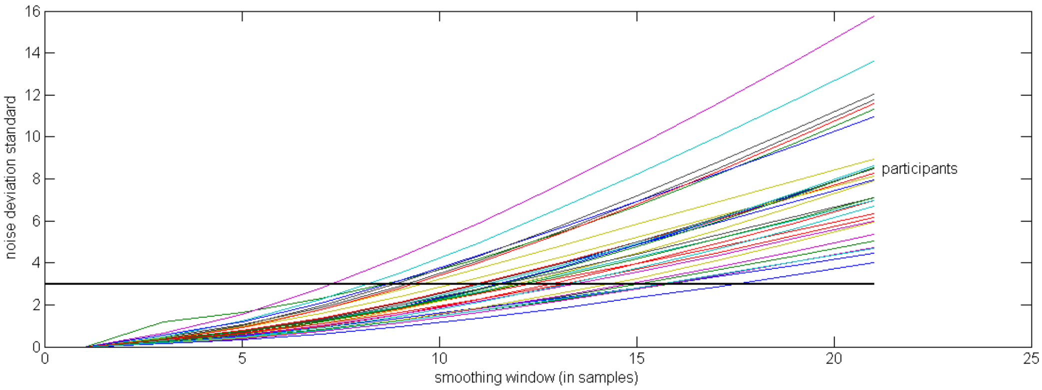


Figure 3.3: Standard deviation vs window size curves

I use the same method both for the listeners and the conductors data: at this point I can finally calculate speed and acceleration that I will use in my analysis.

4. Speed and acceleration

With the 3D positional data at disposal, I can finally build the speed and acceleration signals that will be used in the analysis. To perform the derivative I use the second order Savitzky-Golay filter on each dimension of the positional data: I then obtain the speed signal by calculating the norm among the three axes.

$$speed_x = filter(sig_x); \qquad speed_y = filter(sig_y); \qquad speed_z = filter(sig_z)$$

$$\mathbf{speed}(t) = \sqrt{speed_x^2(t) + speed_y^2(t) + speed_z^2(t)}$$

The filter has 36 taps that correspond with a delay of 0.175 seconds: this parameter is decided, as explained by Amelynck in [4], by inspecting participants movements.

As shown in Figure 4.1 the useful frequency band of the listeners movements is approximately $0 - 4Hz$ that corresponds with the regression window of 0.175 seconds of the filter. The number of taps used is then calculated using the following equation and solving for N (considering $F_s = 100Hz$ and $delay = 0.175s$):

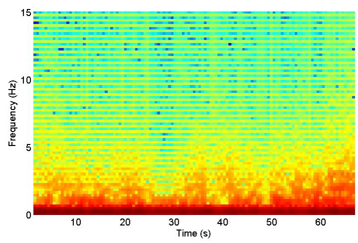
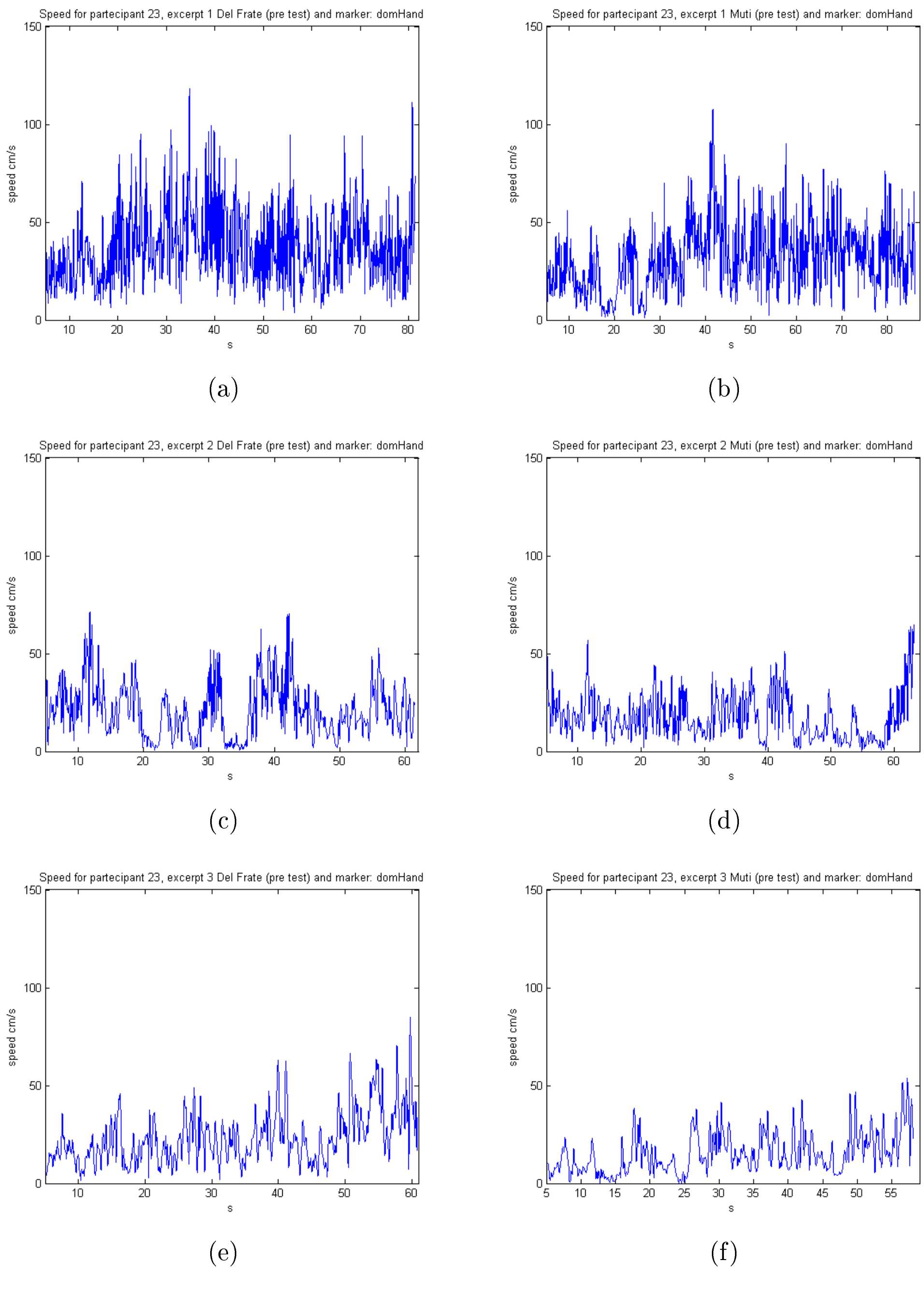


Figure 4.1: Spectrogram

$$\frac{N-1}{2F_s} = delay$$



(a) (b)

(c) (d)

(e) (f)

Figure 4.2: Example of speed plots for participant 23: (a), (c) and (e) represent excerpt DF1, DF2 and DF3; (b), (d) and (f) represent M1, M2 and M3.

I obtain signals as illustrated in Figure 4.2.
When the analysis requires a direct confrontation between participant and conductor movement I have to solve a practical problem: in fact having at disposal also the video of the conductors, I could notice that not all their movements are expressive or related to the act of conducting at all.

(a) (b)

Figure 4.3: Non-expressive gestures

The real issue is that the "bad movement" happens with the *wand* for Muti and with the *hand* for Del Frate. For this reason I have decided to use shorter segments of the excerpts: starting from the same bar (see Fig. 1.2), I take only 2000 samples (20 seconds) that contain only expressive movements. The starting point is decided upon visual inspection on the music excerpt: I have tried to choose segments that contain the most differences possible between the two executions.

		Samples	
Excerpt	**Starting bar**	*Del Frate*	*Muti*
1	5	1945	2090
2	10	3985	4115
3	2	1235	1175

Table 4.1: Choices of starting points

For some applications I will need a rougher representation of the signal, that are only the main features of the movement. For this reason I calculate also the envelope of the signal using again a Moving Average filter with a window size of 250 samples (approximately the length of a bar) and of 500 samples (see Fig. 4.4).

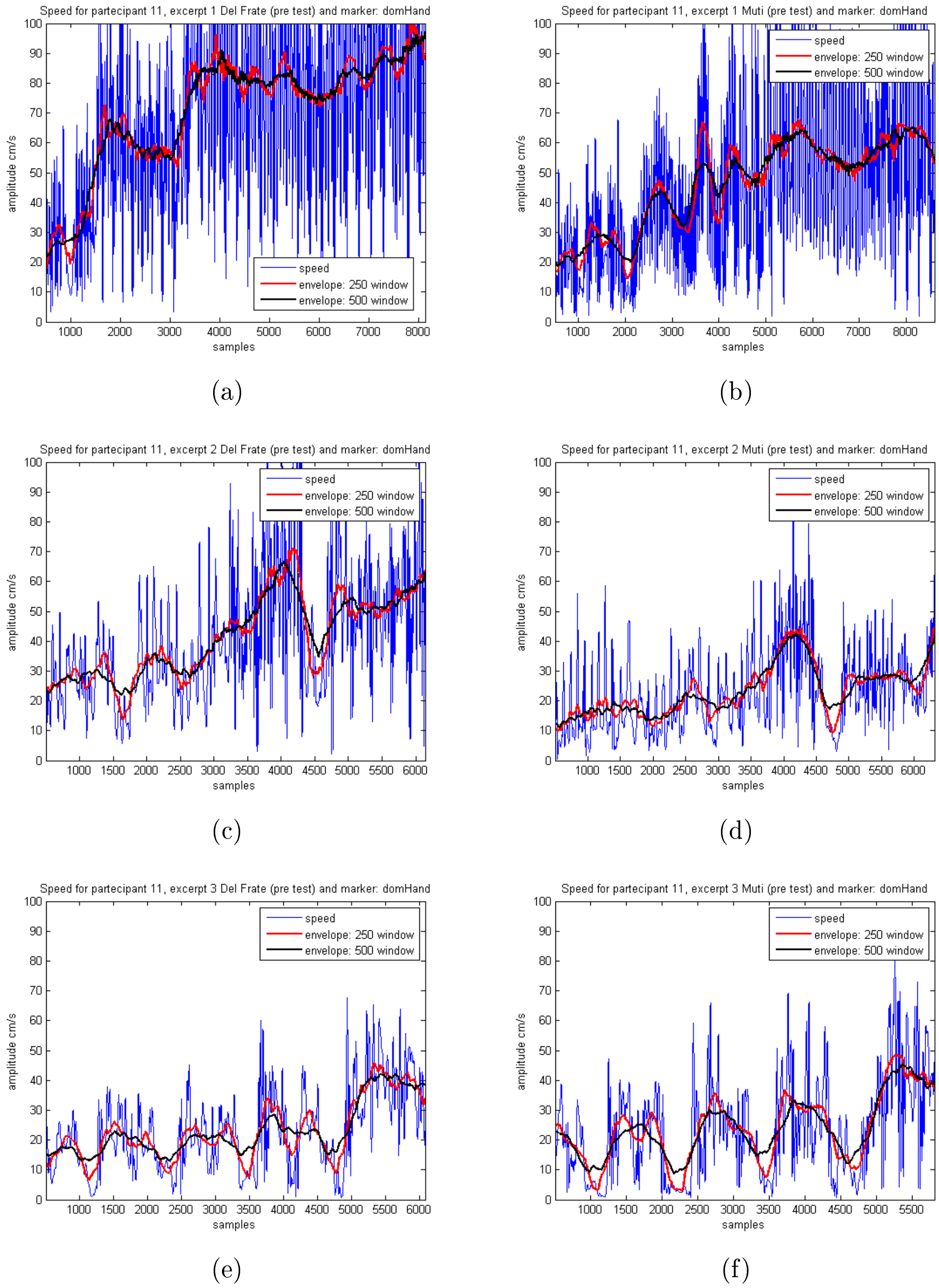
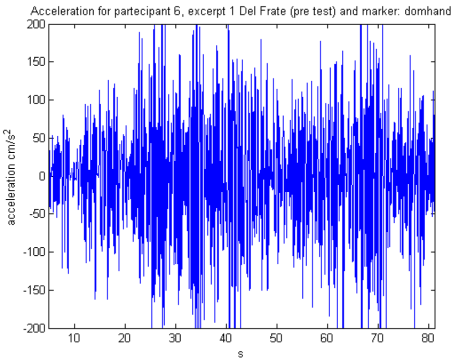
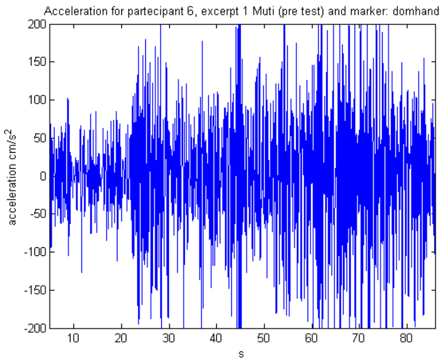


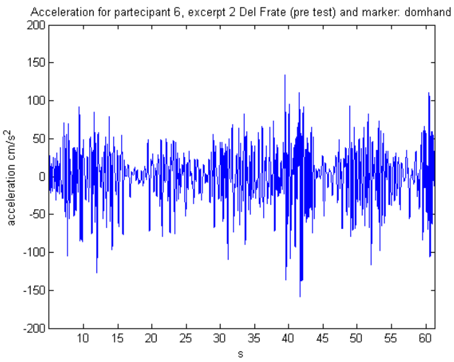
Figure 4.4: Example of speed envelopes for participant 11: (a), (c) and (e) represent excerpt DF1, DF2 and DF3; (b), (d) and (f) represent M1, M2 and M3.



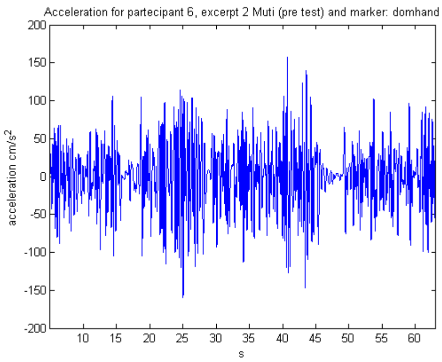
(a)



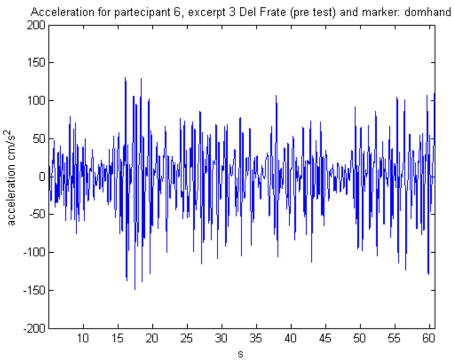
(b)



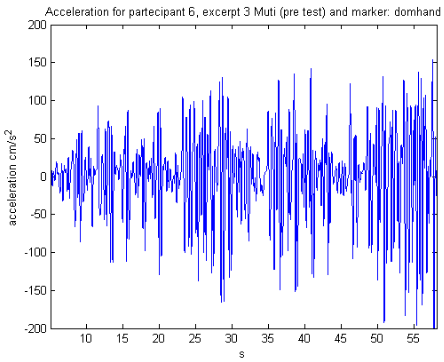
(c)



(d)



(e)



(f)

Figure 4.5: Example of acceleration for participant 6: (a), (c) and (e) represent excerpt DF1, DF2 and DF3; (b), (d) and (f) represent M1, M2 and M3.

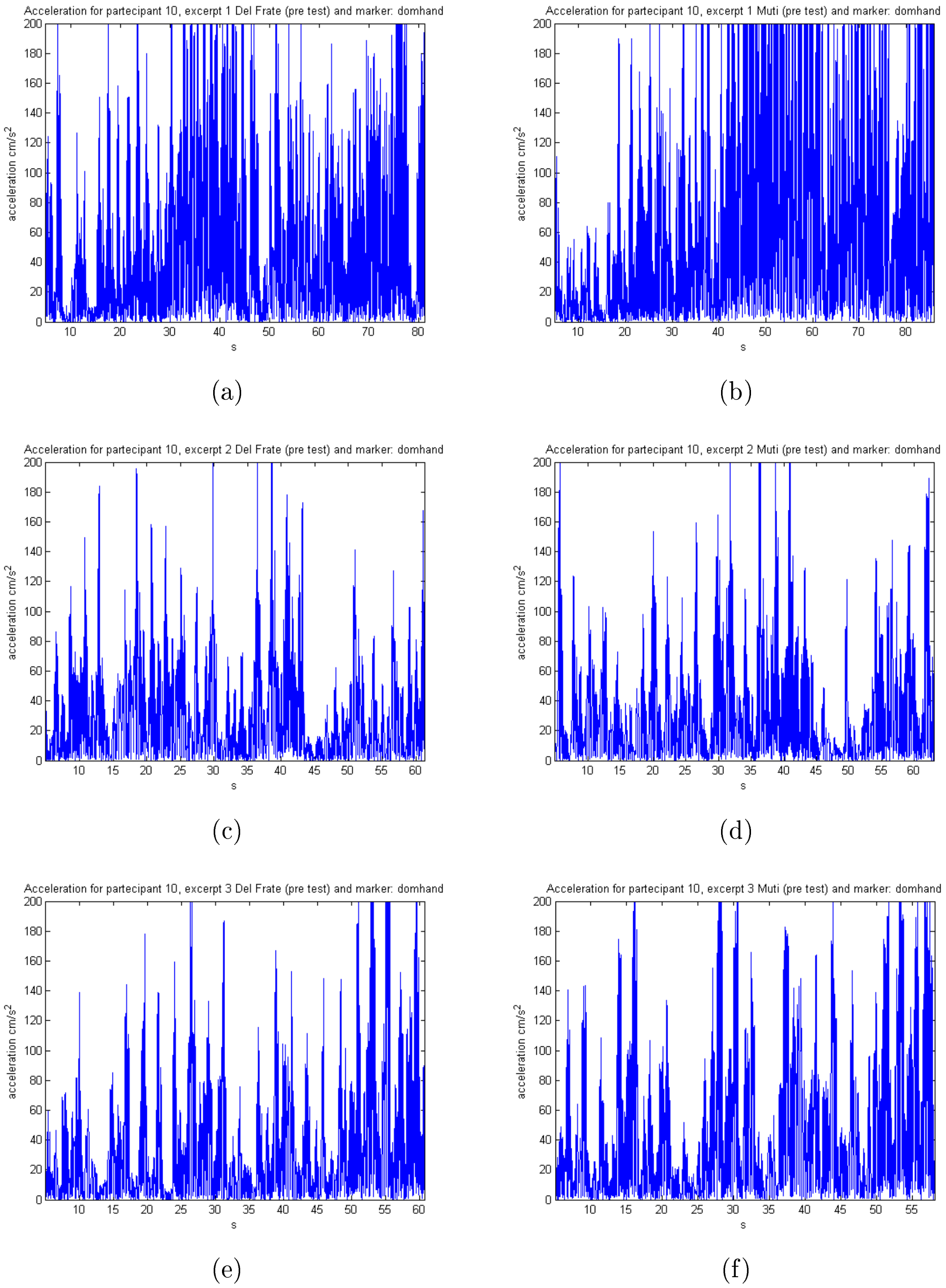
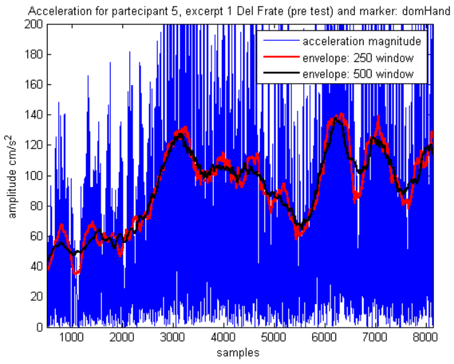
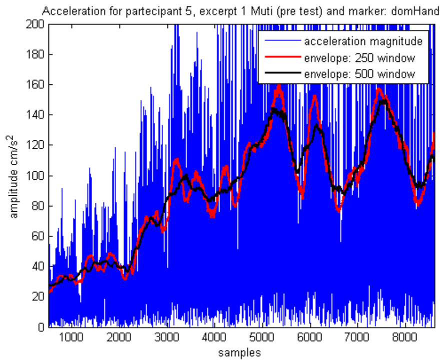


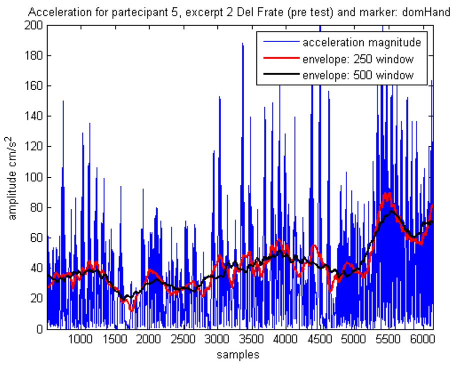
Figure 4.6: Example of acceleration magnitude for participant 10: (a), (c) and (e) represent excerpt DF1, DF2 and DF3; (b), (d) and (f) represent M1, M2 and M3.



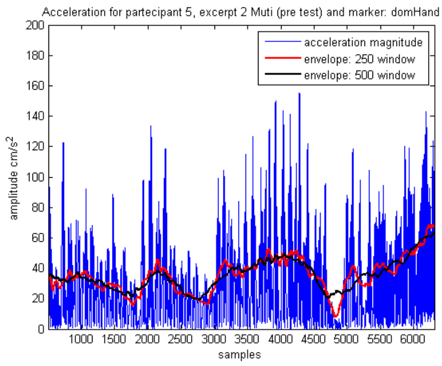
(a)



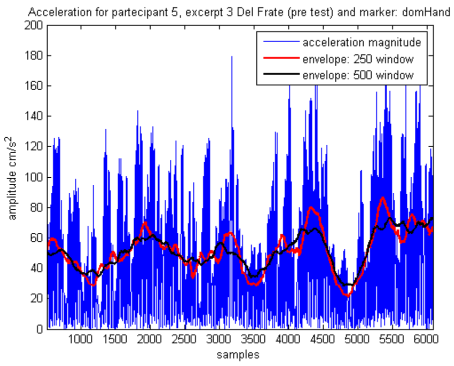
(b)



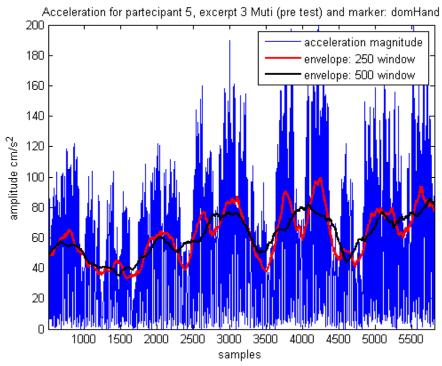
(c)



(d)



(e)

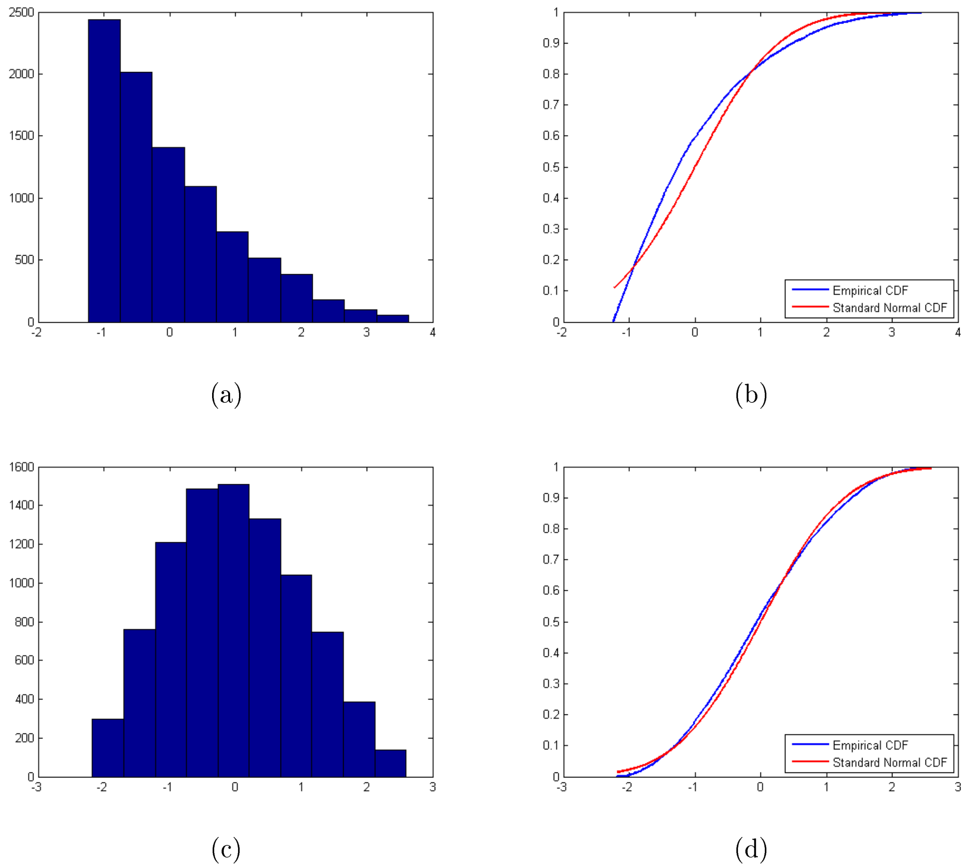


(f)

Figure 4.7: Example of acceleration magnitude envelopes for participant 5: (a), (c) and (e) represent excerpt DF1, DF2 and DF3; (b), (d) and (f) represent M1, M2 and M3.

To obtain the acceleration signal, I filter the speed signal again with the same derivative filter (Savitsky-Golay) and then I take the absolute value to have its magnitude (see Fig. 4.6 and 4.5). With the same procedure as the speed I calculate also the acceleration magnitude envelope with the Moving Average filter (see Fig. 4.7).

Another important step to consider before starting the analysis is to check the signals for normality. For the acceleration *e.g.*, if we look at the histogram or we compare the data with the CDF of a normal r.v. we can see that we don't have a correspondence. As suggested by Desmet in [3], I proceed with taking the square root of the magnitude (Fig. 4.8).



(a) (b) (c) (d)

Figure 4.8: Histogram and CDF before (a-b) and after (c-d) taking the square root of the acceleration magnitude

For the speed envelope in particular we can follow the procedure of Amelynck in [4]: by fitting a Weibull distribution I can assert that the best approximation (see parameters from [5]) is a log-normal distribution. For this reason I will use the *log-envelope* of the speed.

	β	
Excerpt	*Muti*	*Del Frate*
1	2.8062	2.6542
2	1.7619	1.6112
3	1.4237	1.7939

Table 4.2: β paramenter of the Weibull distribution for all excerpts

Part II

Analysis

5. Functional Principal Components Analysis

Theoretic background

Principal Components Analysis (PCA) is widely used in data analysis since it allows to reshape a potentially infinite dimensional problem to a finite one. In the traditional approach a finite dimensional parametric model is used, but if the data are functional a more *ad hoc* methodology is needed. My goal is to show the modes of variation of the data and this is achievable through the study of the *eigenfunction* associated with each *eigenvalue*. We know in fact that the *eigenvalues* of the bivariate variance-covariance function are indicators of the importance of the principal components: in other words, by observing the *eigenvalues*, we can determine how many components are required to have a quality representation of the data. The use of the covariance function (5.1) instead of the correlation function is explained by the fact that, when data are functional, values of the observations $x_i(s)$ and $x_i(t)$, at different times s and t, have the same origin and scale.

$$\nu(s,t) = \frac{1}{N-1}\sum_i \left[x_i(s) - \bar{x}(s)\right]\left[x_i(t) - \bar{x}(t)\right] \tag{5.1}$$

where N is the number of observations and $\bar{x}(k)$ is the mean value at time k among all of them.

My aim is thus to find a weight function ξ that maximize the variation of the probe scores ρ_ξ (5.2)

$$\rho_\xi(x_i) = \int \xi(t)x_i(t)dt \tag{5.2}$$

under the restriction that $\int \xi^2(t)dt = 1$. In other words we want to calculate the variance μ as

$$\begin{aligned}\mu &= \max\left\{Var\left[\int \xi(t)(x_i(t)-\bar{x}(t))^2 dt\right]\right\} \\ &= \max_{\xi}\left\{\sum_i \rho_\xi^2(x_i)\right\} \qquad \text{subject to } \int \xi^2(t)dt = 1\end{aligned} \tag{5.3}$$

where the mean has been removed because it is a well-known variation shared by most of the observations. Because the probe ρ_ξ, being a variably weighted linear combination of function values, is a tool for highlighting specific variation, by maximizing its variance, we want in some way to isolate some trends. We call μ and ξ the largest *eigenvalue* and *eigenfunction*.

It is important to notice that in functional data the number of values n is usually much greater than the number of observations N: this implies that the maximum number of non-zero *eigenvalues* is $\min\{N-1, n\}$ and in most of the cases this is $N-1$. So for each choice of $\ell \in [1, N-1]$ the ℓ principal *eigenfunctions* define an orthogonal basis system ξ_ℓ that can be used to approximate the sample function x_i. An important characteristic of the basis ξ_ℓ is that it is the most efficient possible among the bases of size ℓ in the sense that the total error sum of squares is the minimum possible (5.4).

$$\ell = argmin\left\{\sum_i^N \int [x_i(t)-\bar{x}(t)-\mathbf{c}_i'\xi_\ell(t)]^2 dt\right\} \tag{5.4}$$

The number of ℓ bases that have to be used is decided upon a visual inspection of a plot of the *eigenvalues* μ_j versus the indices j: the optimal total square error is in fact equal to the sum of the discarded *eigenvalues*. We will therefore choose the number of bases ℓ equal to the index that minimizes the sum (5.5). This kind of plot is usually referred as *scree plot*.

$$\ell = argmin\left\{\sum_{j=\ell+1}^{N-1} \mu_j\right\} \tag{5.5}$$

The coefficient vector $\mathbf{c}_i$ in (5.4) describes the optimal fit to each function x_i and its elements are called *principal components scores*: we will use them to interpret the variation identified by the PCA.

$$c_{ij} = \rho_{\xi_j}(x_i - \bar{x}) = \int \xi_j\left[x_i(t)-\bar{x}(t)\right] dt \tag{5.6}$$

Application

Practically, what I'm trying to do, is express each subject signal (square root of acceleration magnitude) as a sum of fixed functions. To perform the analysis I use the *FDA Matlab Toolbox* of Ramsay (see [6] and [7]).

$$\tilde{f}(t) = \bar{f}(t) + \sum_{k=1}^{\ell} c_{ik}\xi_k(t) \tag{5.7}$$

where

- $\tilde{f}(t) \rightarrow$ is the performance of each subject $(i = 1, ..., 32)$
- $\bar{f}(t) \rightarrow$ is the mean performance among all participants (**commonality**)
- $\ell \rightarrow$ is the number of eigenfunctions chosen
- $c_{ik} \rightarrow$ component scores, factor that weights the eigenfunctions (**individuality**)
- $\xi_k(t) \rightarrow$ eigenfunctions (**commonality**)

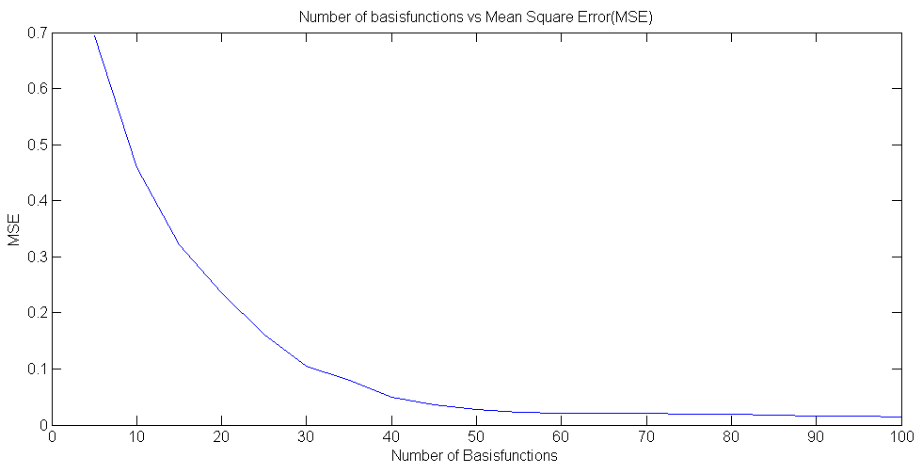


Figure 5.1: MSE plot

The number of basis-functions used is determined upon inspection of plots as Figure 5.1: here it is presented the variation of the MSE error (confrontation between original signal and approximated one) with the increasing of the number of basis-functions. Looking at all the possibilities, 60 basis-functions are a good trade-off; but working with less functions could reduce the computational costs (calculating eigenfunctions has a complexity of $O(K^2)$ with

K the number of functions): for this reason I choose to calculate the number of eigenfunctions K that explain at least 70% of variability among the participants (see Fig. 5.2).
For my experiment I verified that for all excerpts 3 eigenfunctions are always enough to cover at least this percentage of variability and in most of the cases this is around $80 - 90\%$.

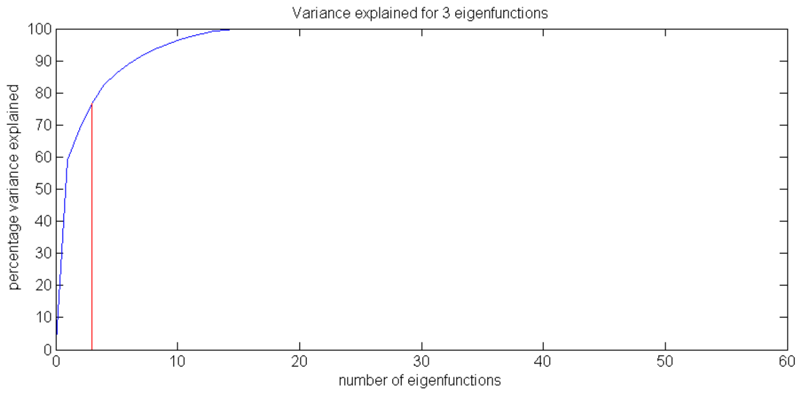


Figure 5.2: Example for excerpt M2 of variability covered with 3 eigenfunctions

The analysis is performed distinctly for each excerpt (pre- and post-test) and for all the participants together. In Figures 5.3-5.8 the eigenfunctions are presented as variation in respect to the mean (blue curve): the red and the green curves represent plus and minus the largest eigenvalue among all subjects. The gap between red and green curve is proportional to the amount of variance explained. At the end of the analysis each participant is represented by a vector containing the component scores and this, together with the set of eigenfunctions, is enough to identify his/her movement.

Excerpt 1 (Fig. 5.3-5.4)

For both tests the first eigenfunction covers most of the variability in the whole segment (we can see that the curves never intersect): in particular for the post-test the only first eigenfunction would be enough (more than 70% of variability). This can be explained by the fact that the movement of the group tends to have the same shape, differing only in amplitude.

It's interesting to notice that both for the Muti and the Del Frate segments, in the second eigenfunction the excerpt is split in half: in the first part the green curve is above the red one and in the second half the opposite. The subjects that have an high coefficient for this eigenfunction will have a movement higher than the average for the first part and lower for the second.

Excerpt 2 (Fig. 5.5-5.6)

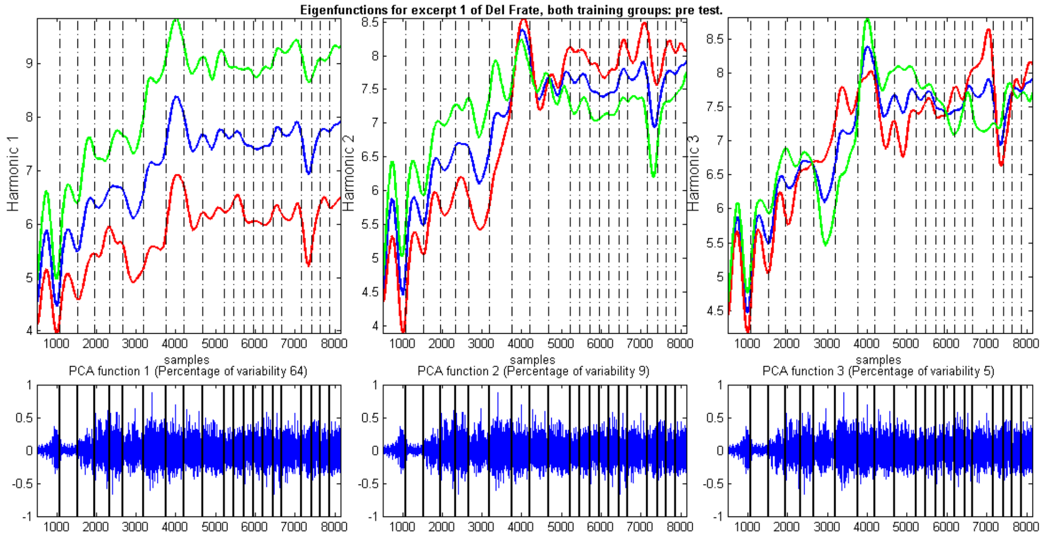
Also for the second excerpt we can see that the variability explained increases between pre- and post-test, especially the one relative to the first eigenfunction: for both excerpts in the post-test the first eigenfunction explains the variability of the whole segment (the curves do not intersect), meaning that all the participants tend to move in the same way.

In the post-test one eigenfunction already covers more than 70% of variability: while in the pre-test the third eigenfunction contains important and meaningful parts of the movement, in the post-test it is responsible only of small details (small gap between green and red curve).

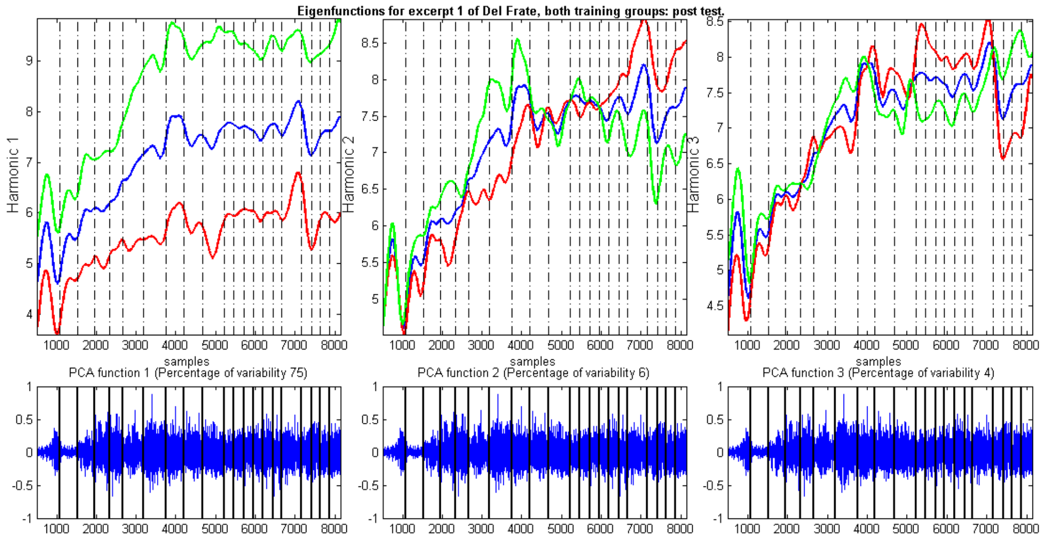
Excerpt 3 (Fig. 5.7-5.8)

For the third excerpt this behavior is less visible: while in the post-test the curves do not intersect, however the variability seems to diminish. It is therefore interesting in this case to perform the analysis distinctly for the two training groups (see Fig. 5.9): for the trained excerpts the variability increases in both cases but we need two or all three eigenfunctions to reach the desired variability covered.

Figures

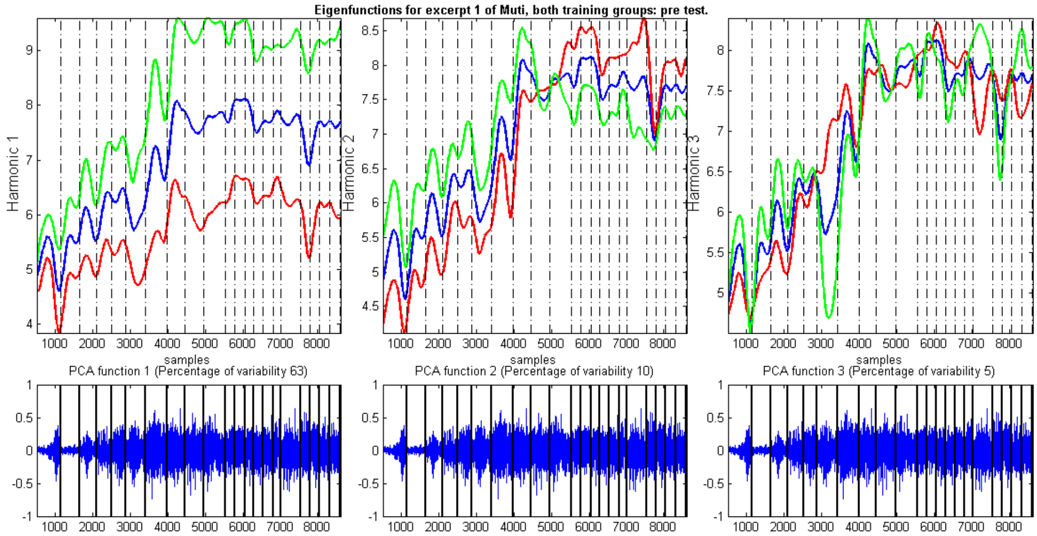


(a)

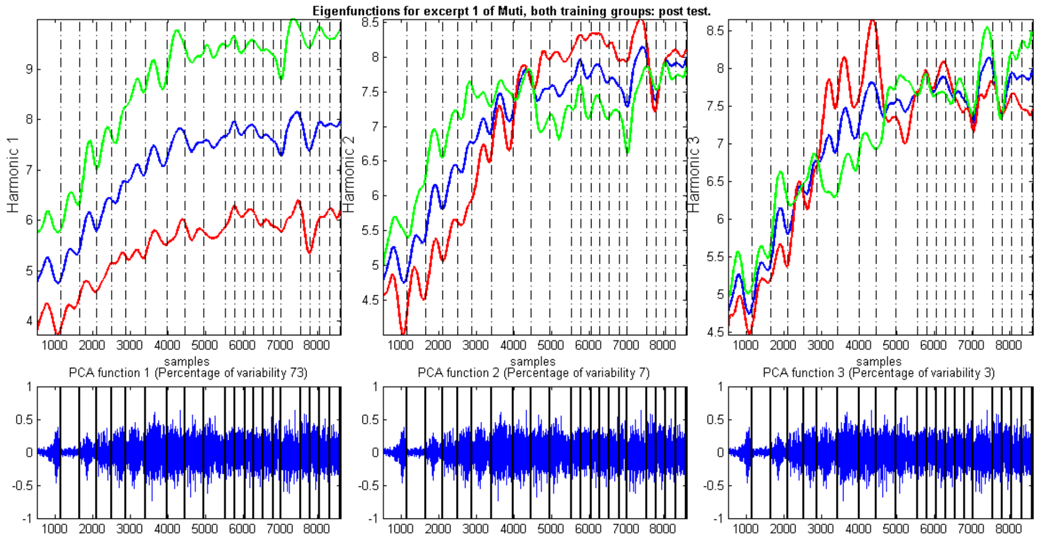


(b)

Figure 5.3: First three eigefunctions : excerpt DF1 pre-test (a) and post-test (b)

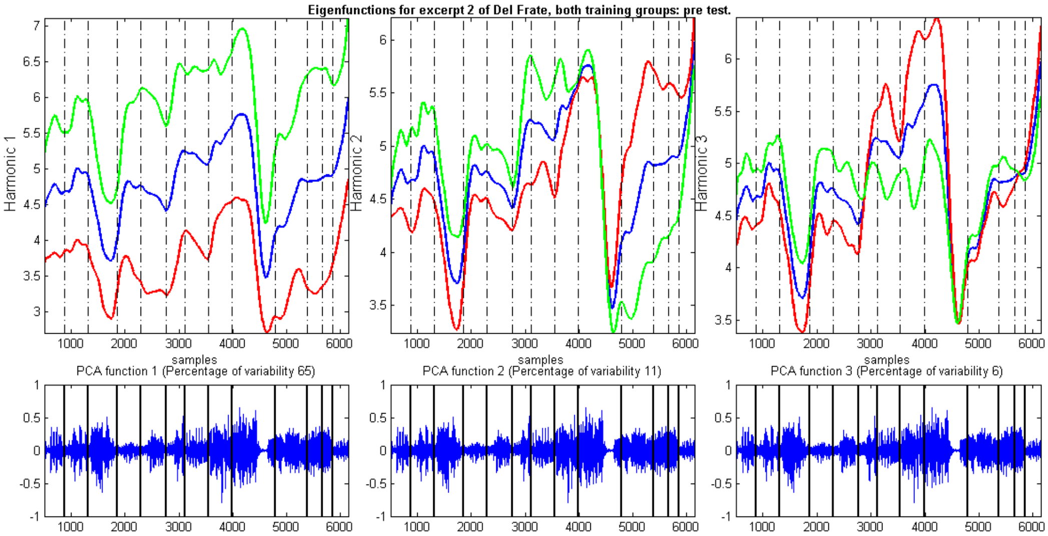


(a)

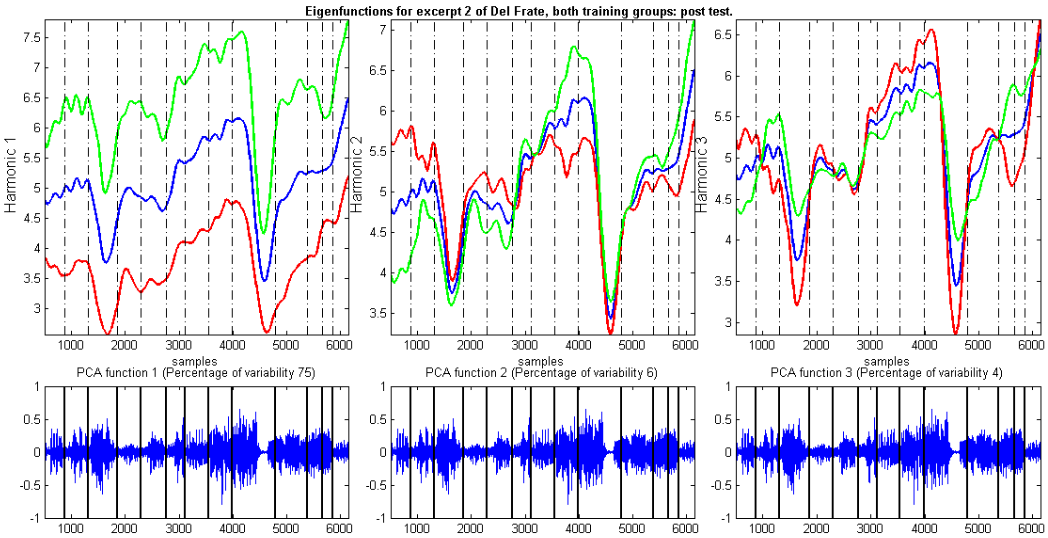


(b)

Figure 5.4: First three eigefunctions: excerpt M1 pre-test (a) and post-test (b)

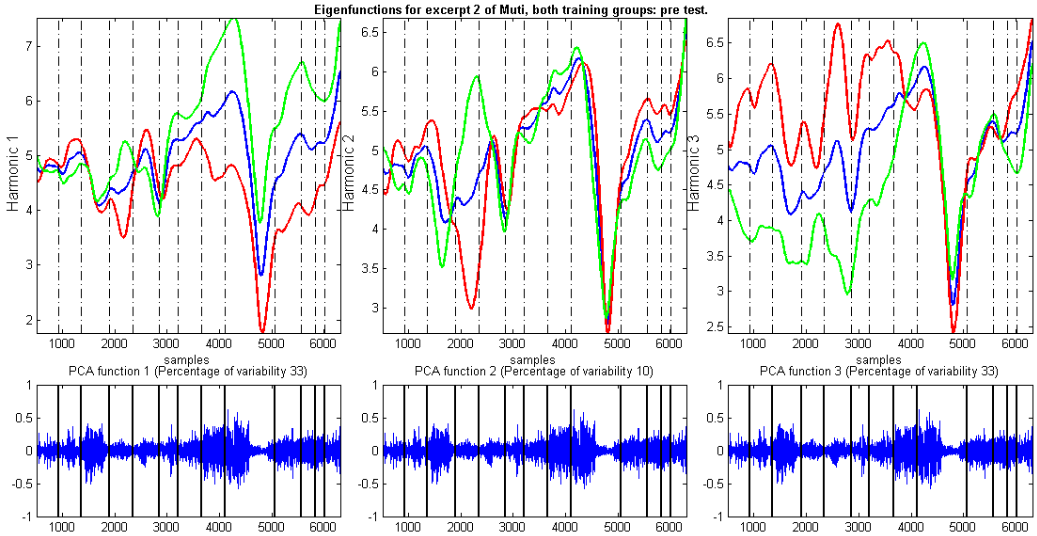


(a)

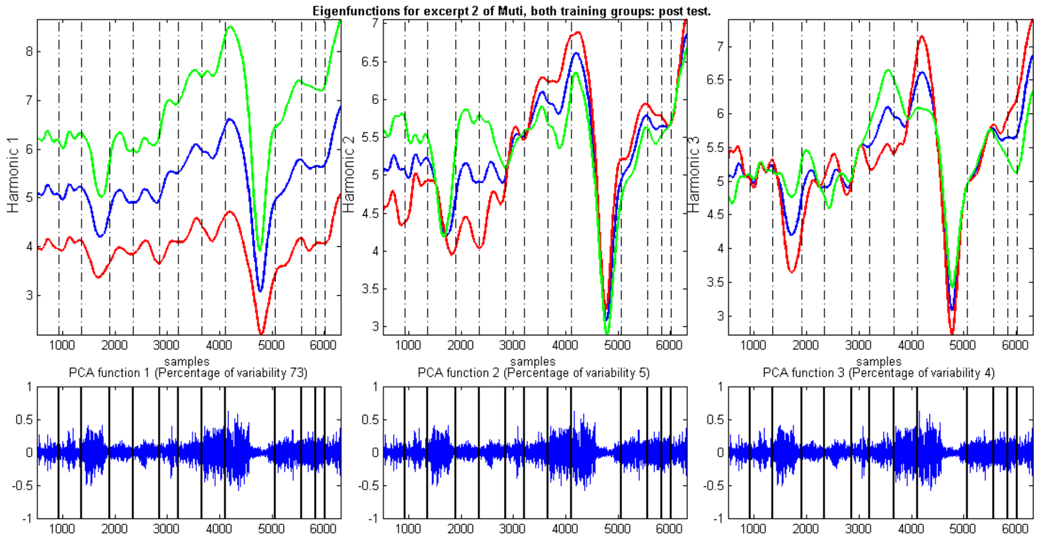


(b)

Figure 5.5: First three eigefunctions: excerpt DF2 pre-test (a) and post-test (b)

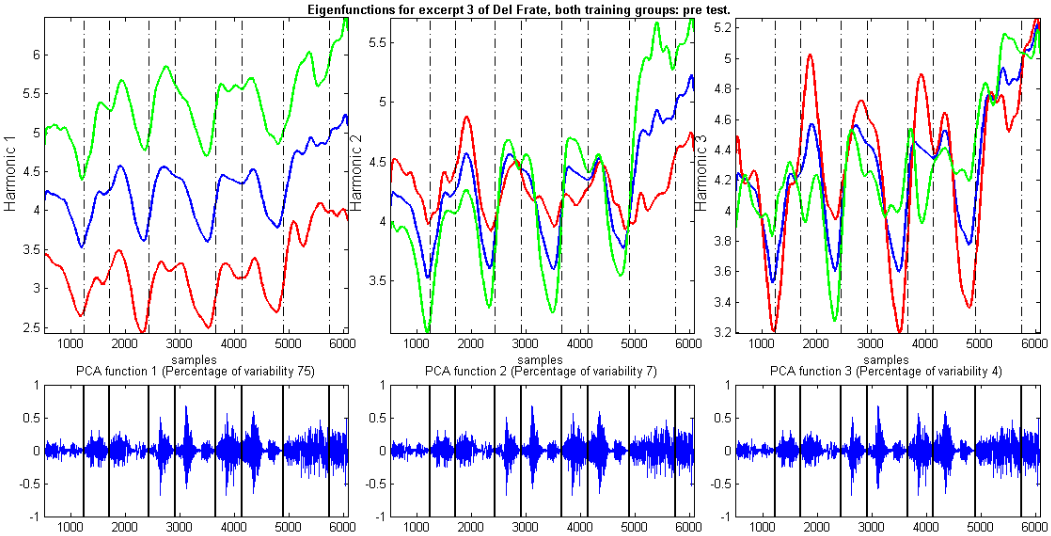


(a)

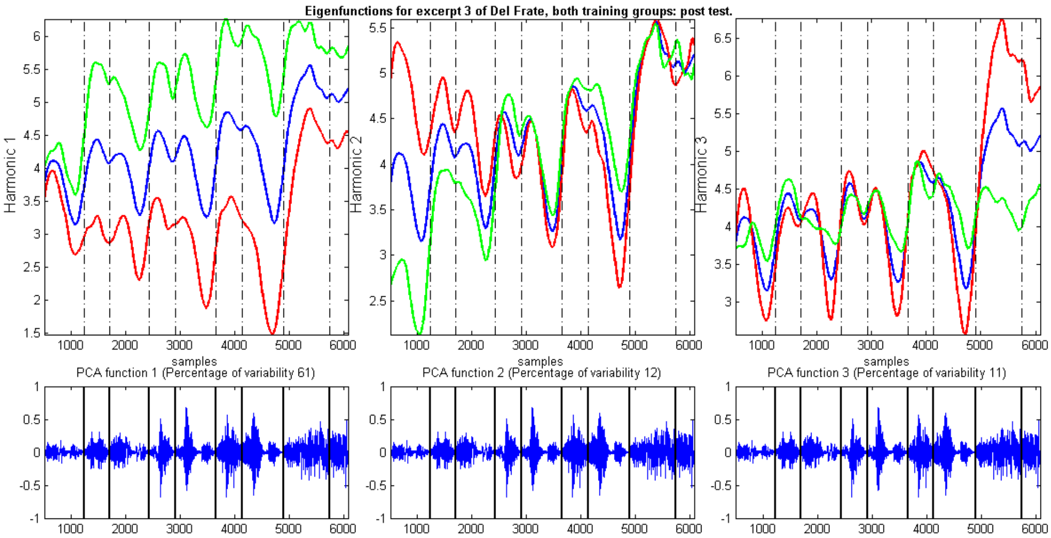


(b)

Figure 5.6: First three eigefunctions: excerpt M2 pre-test (a) and post-test (b)

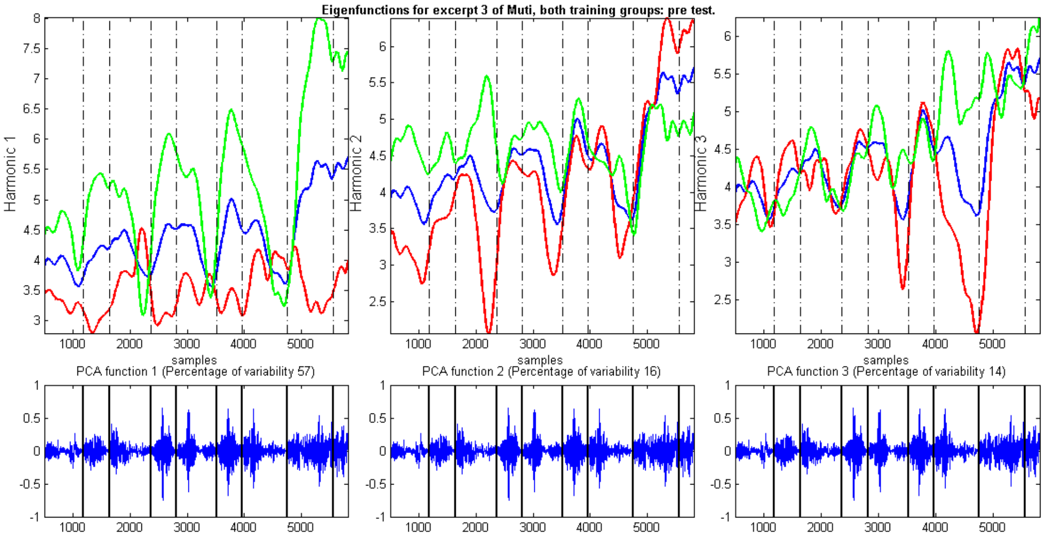


(a)

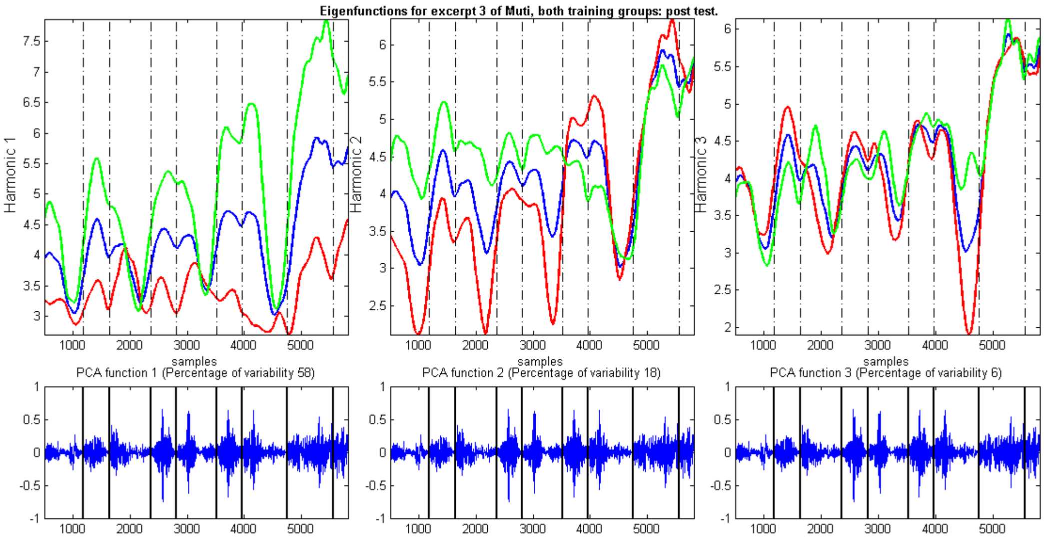


(b)

Figure 5.7: First three eigefunctions: excerpt DF3 pre-test (a) and post-test (b)

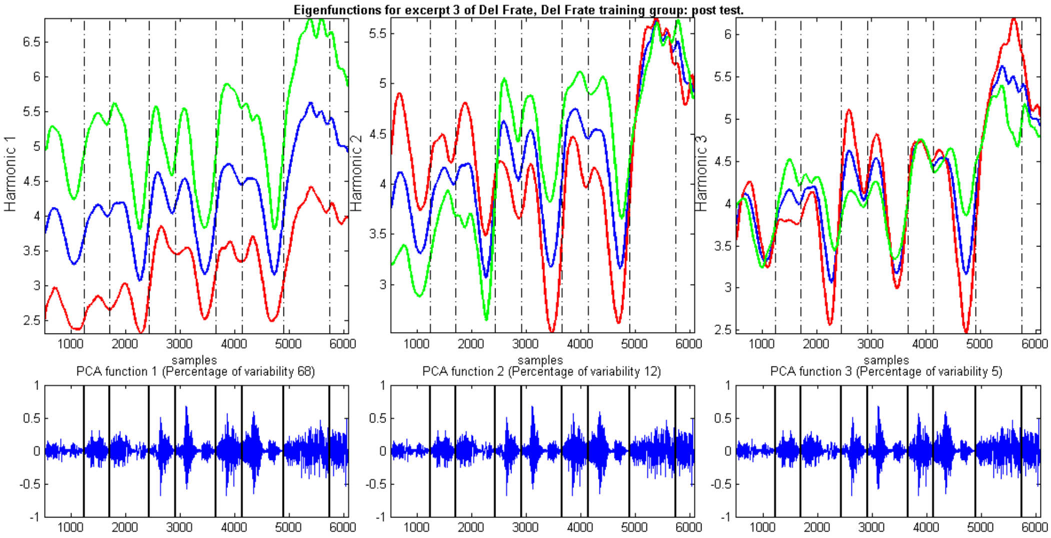


(a)

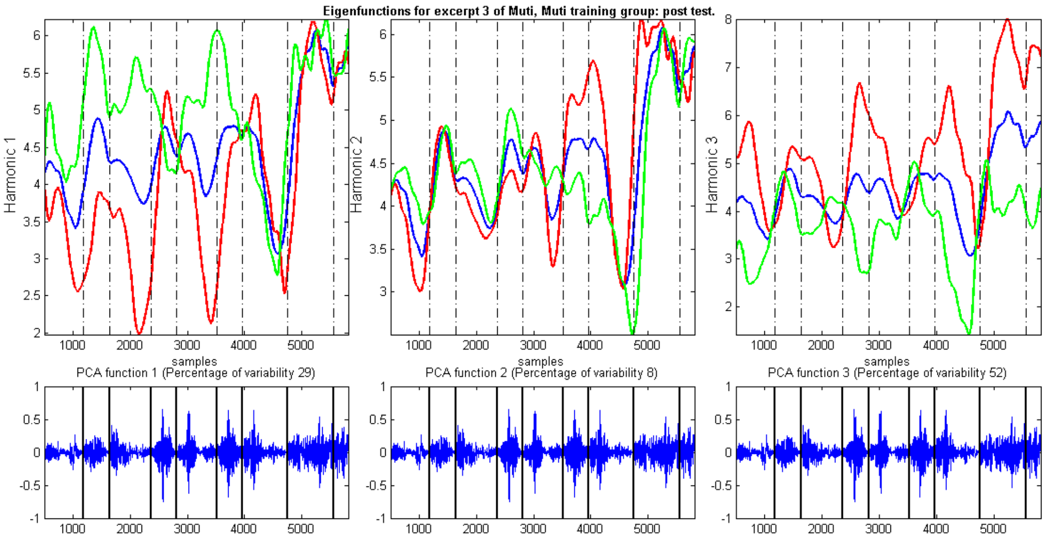


(b)

Figure 5.8: First three eigefunctions: excerpt M3 pre-test (a) and post-test (b)



(a)



(b)

Figure 5.9: First three eigefunctions of Del Frate training group excerpt DF3 post-test (a) and Muti training group excerpt M3 post-test (b)

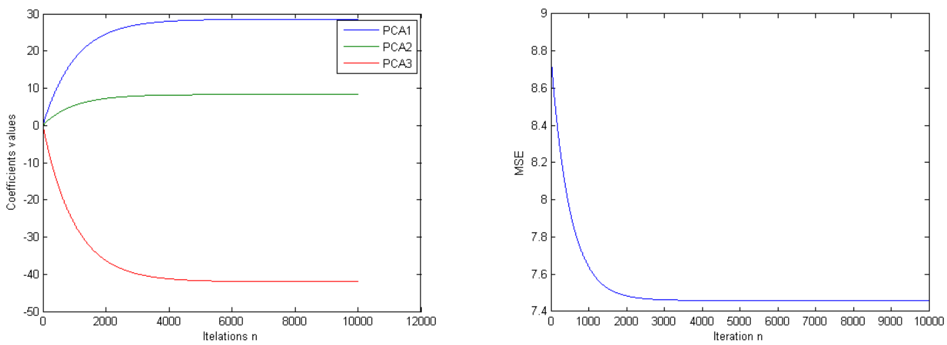
Conductors movements

To compare the conductors movement with the participants ones it is necessary to have also for the conductor a vector of component scores derived from the same set of eigenfunctions. For this reason I have decided to implement Least Mean Square (LMS) algorithm (Listing 5.1).

```
iter=1e4;
C=zeros(iter ,length(eigftoshow));
mu=0.001;
%x=eigenfunctions
%fun=original conductor signal (without mean)
for n=1:iter

        y=C(n,:)*x;
        er=fun-y;
        C(n+1,:)=C(n,:)+(mu*er*conj(x).');

        err(n)=mean(er.^2);

end
```

Listing 5.1: Code for LMS

As we can see from Figure 5.10 both the coefficients and the error converge quite fast but it is evident that the error remains too high. If we have a look at the reconstructed signal (Fig. 5.11) we can see that three eigenfunctions don't express even the main features of the movement: this can tell us already that the participant movements and the conductor ones are totally different and we need a more indirect method to confront them.



(a) (b)

Figure 5.10: Convergence of coefficients (a) and error (b) during LMS algorithm

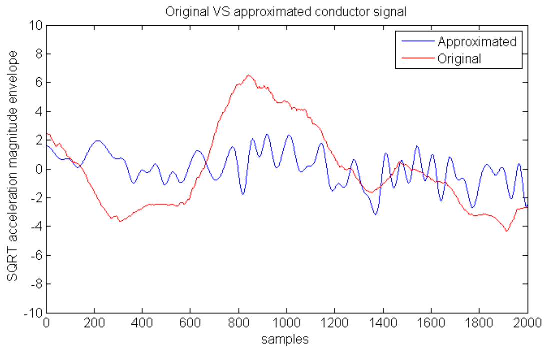


Figure 5.11: Original VS approximated conductor signal with 3 eigenfunctions

One important thing to notice at this point is that, while 3 eigenfunctions are enough to describe all the variability among the participants, we need more of them to be able to reconstruct the conductor signal. Trying with 16 eigenfunctions the differences are notable (see Fig. 5.12).

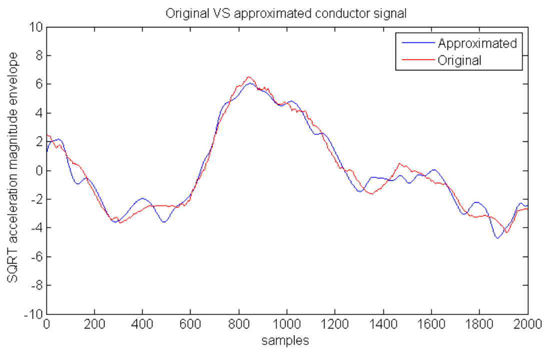


Figure 5.12: Original VS approximated conductor signal with 16 eigenfunctions

Because the conductor signal is derived from the participants movement, I try now to check the variation of the MSE (between approximated conductor signal and original one) with the increasing of the number of eigenfunctions used, setting up the analysis separately for the two training groups. For example for excerpt 1 in Figure 5.13, we can observe that while for the trained segment I have a reduction of the error between pre- and post-test, for the untrained one there is an increase. It seems that the training had a positive influence on the trained segments but a negative one on the untrained.

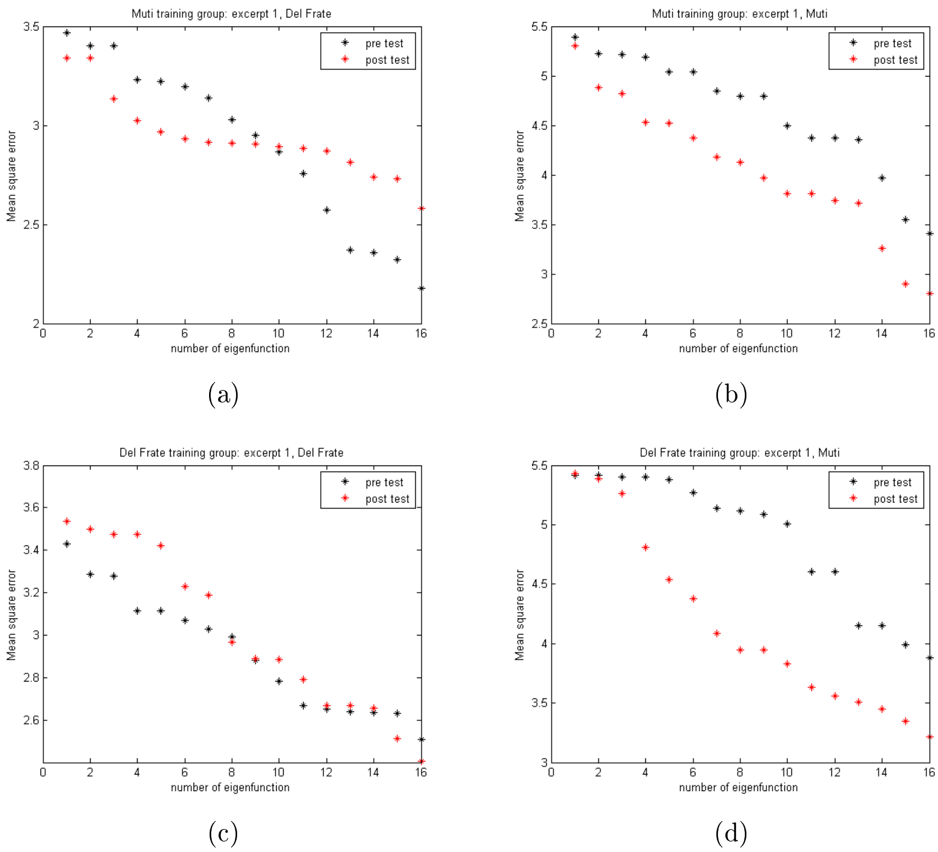


Figure 5.13: Plots of MSE varying the number of eigenfunctions

6. Direct inspection

While for the PCA analysis I used the acceleration signal, I want now to inspect the speed of the participants: in particular I will use the mean speed among all the participants. As it is evident in the example of Figure 6.1, the speed envelope of the subjects has almost the same shape as the music amplitude except for a scaling value.

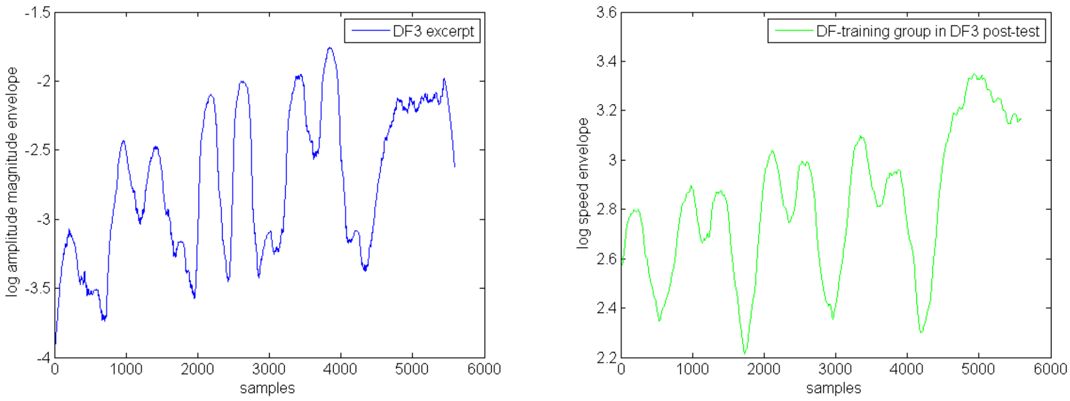


Figure 6.1: Example of confrontation between log of music amplitude magnitude (left) and log of speed envelope (right).

For this reason I tried to look directly at the plots of the speed envelope signals trying to find any correspondence with the music amplitude. The most interesting excerpt is the third one, that is the one with the most differences between the two executions. In fact as I explained in Chapter 1 (pg. 6), beside the tempo differences, this excerpt has also a discrepancy in the amplitude. I want to see if I can find this trend also in the movement.

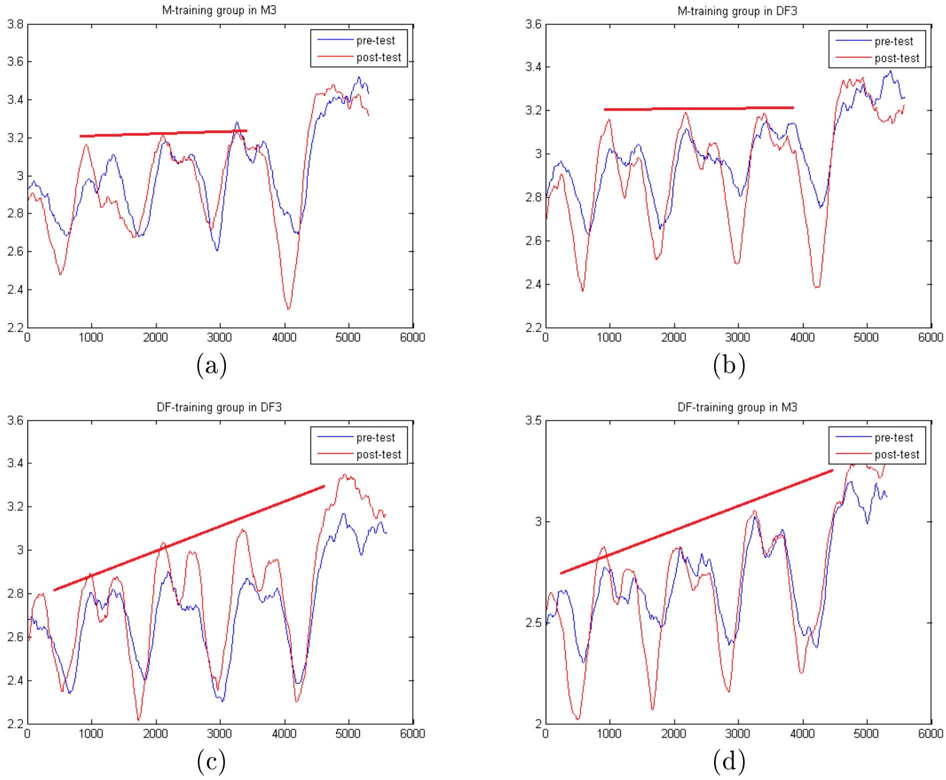
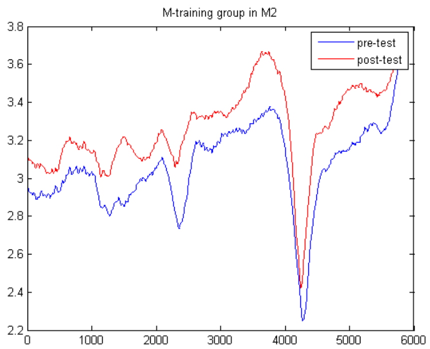


Figure 6.2: Plots of speed envelope signals in pre- and post-test for excerpt 3: (a)-(b) Muti training group and (c)-(d) Del Frate training group.

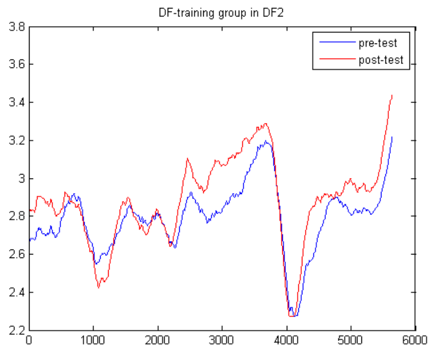
As we recall from Figure 1.3 the repetitions in the Del Frate excerpt have an increasing trend in amplitude, while the ones in the Muti excerpt have a more flat one. Now we can find this behavior again in the participants movements (Figure 6.2): for the Muti training group *e.g.*, the flat trend, typical of excerpt M3, is not only present in the speed signal relative to this excerpt, but also in the DF3 one. In the same way the increasing trend of DF3 is mantained from the Del Frate training group both on the trained excerpt and on the un-trained one.

This is a further proof that the training, despite what was my original hypothesis, influences also the un-trained segments, even if in a "negative" way.

Another interesting feature we can notice not only for excerpt 3 in Figure 6.2 but also for excerpt 2 (see Fig. 6.3) is the phenomenon of *anticipation* due to the learning (see [8] and [9]). As we can see, the red curves (relative to the post-test) are always skewed to the left in respect to the blue ones (pre-test): that is because knowing the music, participants tend to anticipate their movements remembering in some way what it is to come. Because they expect the upcoming event they react more quickly: in fact we can see that this anticipation effect is more prominent during "important" events in the music, that is for example big changes on speed.



(a)



(b)

Figure 6.3: Example of anticipation in trained excerpt: M2 (a) and DF2 (b).

7. Correlation analysis

Another interesting way to treat the data is to look at the correlation among participants belonging to the same group, between participants speed movements and music amplitude and between participants and conductors acceleration movements.

Figure 7.1 is an example of correlation between the movement of participants, that is how each sample of all participants correlates with every other sample. If I have a matrix with the various samples on the columns and the various participants on the rows, I am calculating the correlation between each pair of columns of this matrix.

$$\begin{bmatrix} x_{p_1}(1) & \cdots & x_{p_1}(t) & \cdots & x_{p_1}(N_{sam}) \\ \vdots & \ddots & & & \vdots \\ x_{p_k}(1) & & x_{p_k}(t) & & x_{p_k}(N_{sam}) \\ \vdots & & & \ddots & \vdots \\ x_{p_{N_p}}(1) & \cdots & x_{p_{N_p}}(t) & \cdots & x_{p_{N_p}}(N_{sam}) \end{bmatrix}$$

One can see that after the training, the correlation among the participants changes:

- for the Del Frate training group the correlation seems to decrease but there are more defined coherence intervals (the high correlation squares on the diagonal);
- for the Muti training group the correlation gets really higher everywhere but still one can identify the same coherence intervals.

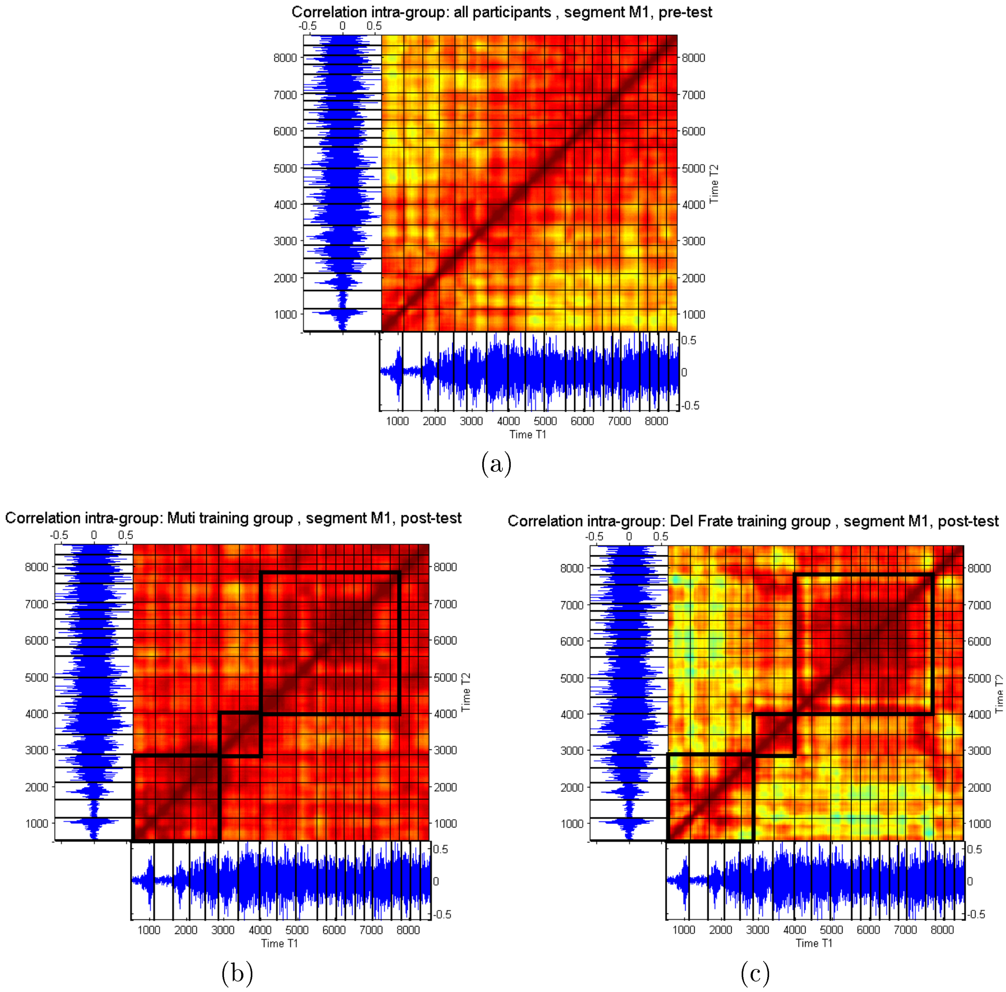


Figure 7.1: Correlation intra-group: pre-test (a), post-test Muti training group (b) and post-test Del Frate training group (c) for excerpt M1. Correlations values are from -1 (blue) to 1 (red).

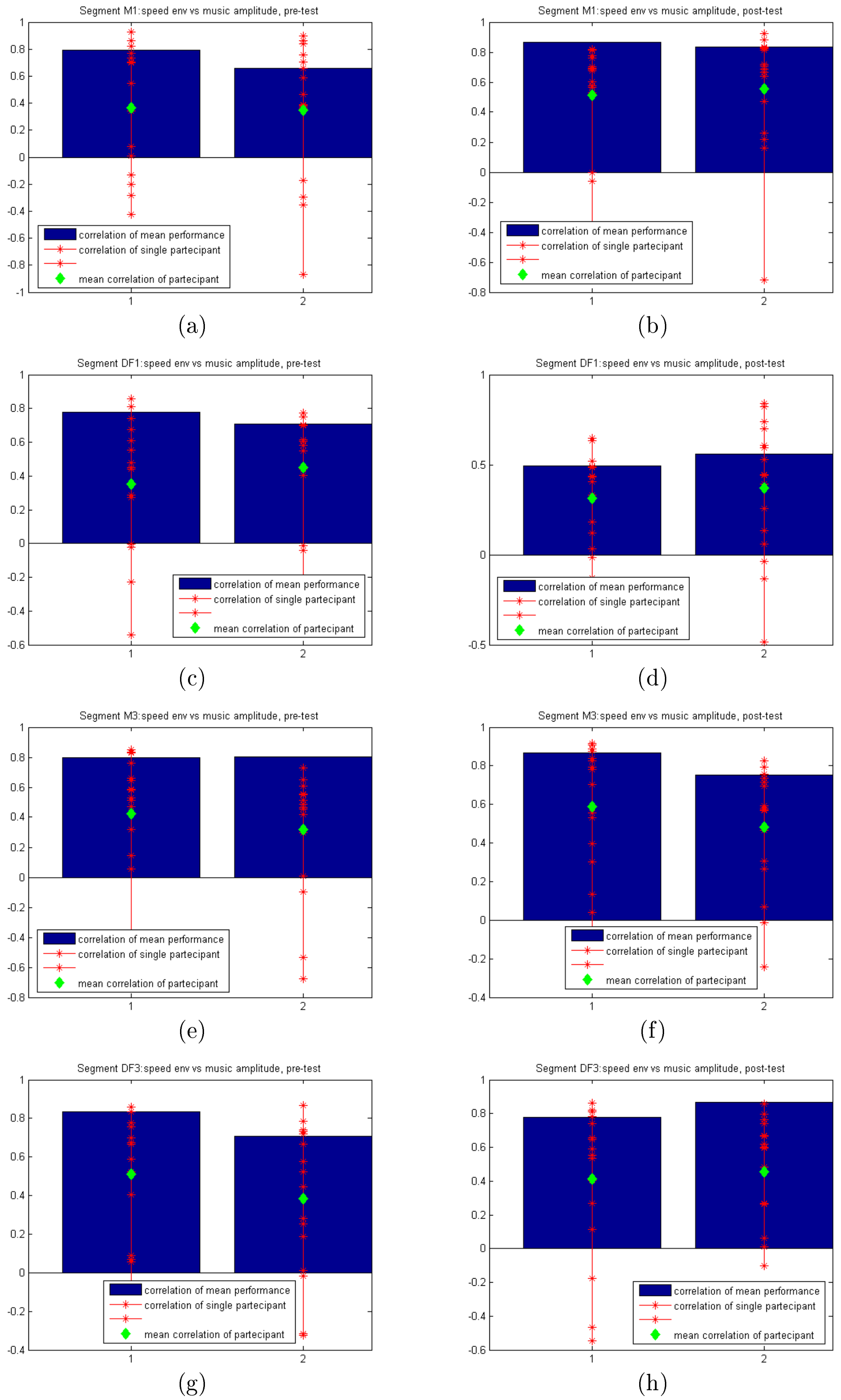


Figure 7.2: Correlation between log speed envelope and music amplitude for excerpt DF1 [(a)-(b)], M1 [(c)-(d)], DF3 [(e)-(f)] and M3 [(g)-(h)]. Pre-test plots are on the left and post-test on the right.

Figure 7.2 shows some examples of the correlation between speed envelope of participants and music excerpts: doing the direct analysis in Chapter 6 in fact I noticed that the speed-movement was very similar in shape to the music amplitude and for this reason I think that it can be interesting to try to confront them.

To perform this analysis first I calculate the mean performance (speed envelope signal) within one training group by taking the mean among subjects for each time sample; then I proceed on doing the correlation (Pearson Linear correlation) between the mean performance and the music amplitude (blue blocks, *1* for Muti training group and *2* for Del Frate's) and between the single participants and the music (red stars). Highlighted in green is the mean of the correlations of the single subjects.

The first interesting thing to notice is that there are cases where the correlation between the mean performance and the music is higher than any other correlation of one single participant (in Figure 7.2 it happens when the blue block is higher than any red star): in other words the mean speed among all subjects in the group can perform better (in term of correlation) than the members of the group themselves.

Furthermore, while for trained excerpt the correlation increases after the training, in some cases it decreases for non-trained one as for example for excerpt M3 and DF3 (Figure 7.2 (e)-(f)):

- for the Muti excerpt we have that for the Muti training group the correlation goes from 0.796 to 0.867 while for the Del Frate training group from 0.803 to 0.754;

- in the same way for the Del Frate excerpt it goes from 0.708 to 0.865 for the Del Frate training group and from 0.833 to 0.777 for the Muti training group.

It seems that training has a sort of negative effect on the untrained excerpt beside the obvious improvement on the trained ones. Because the music is almost the same, subjects tend to concentrate on the known melody more than on the different tempo. The fact that also for the pre-test the correlation is quite high, proves that subjects movement wants to reproduce the music amplitude.

In Figure 7.3 there are the plots of the correlation performed between participants and conductors acceleration magnitude envelope. In this case I used the Kendall correlation that is a type of non-linear correlation that checks the concordance in the variations of the signal. In other words, being $x_1, ..., x_N$ the samples of the mean performance among the group and

$y_1, ..., y_N$ the samples of the conductor movement, one can define the correlation coefficient τ as

$$\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\text{number of pairs}}$$

where a pair (x_i, y_i) is called *concordant* if $\forall$ j the pair (x_j, y_j) has to meet one of the following criteria:

$$\{x_i > x_j \ \& \ y_i > y_j\} \bigwedge \{x_i < x_j \ \& \ y_i < y_j\}$$

All other possibilities lead to a *discordant* pair.

Despite the fact that the correlation is generally lower than the one in Figure 7.2, one can notice the same behavior as before: values increasing for trained excerpts and decreasing for non trained ones.

While for the music amplitude the correlation was high for all excerpts, in this case we have a distinction: in fact it is quite low for excerpt 1 (even negative in the pre-test) that has a faster tempo and slightly higher for excerpt 2 that has a slower tempo.

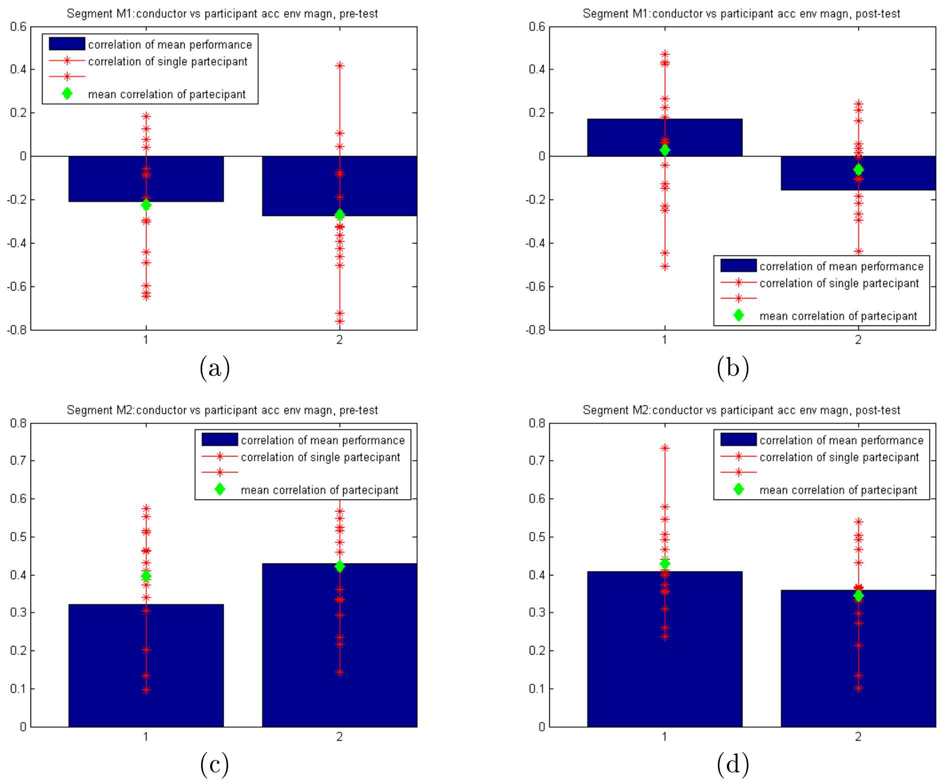


Figure 7.3: Correlation between acceleration magnitude envelope of participants and of conductor for excerpt M1 [(a)-(b)] and M1 [(c)-(d)]. Pre-test plots are on the left and post-test on the right.

8. Mutual Information

Another possible way to check for correspondences between participants and conductors movements (in particular acceleration) is to treat the signals as random variables and look at their mutual information (see [10] and [11]). From Shannon's definition we have

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)}$$

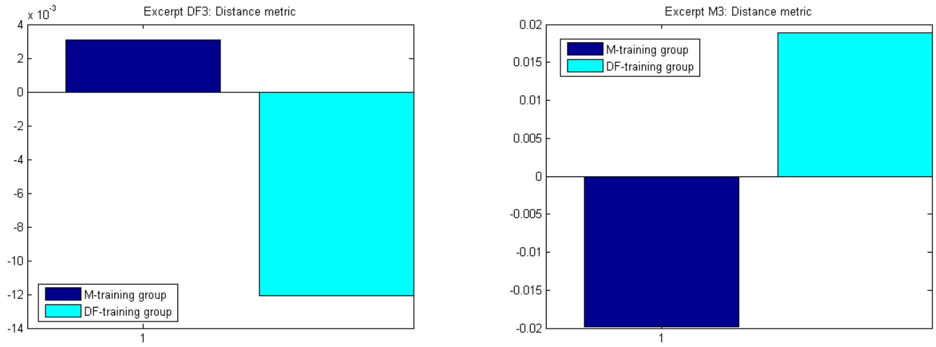
where X and Y are random variables, $p(x,y)$ is the joint probability density function, $p(x)$ and $p(y)$ are the marginal probability density functions for X and Y and $H(\cdot)$ is the entropy.

From mutual information and entropy we can obtain a distance measure that allows to put into one number the differences between conductor and participants gestures.

$$D(X,Y) = 1 - \frac{I(X;Y)}{max\{H(X), H(Y)\}}$$

where we can consider X as the mean acceleration envelope among all participants and Y the acceleration envelope of the conductor.

The results, particularly for excerpt 3, are what we expect as one can see in Figure 8.1 where the difference between distance conductor-participants in post- and in pre-test is plotted: the negative value means an improvement (distance getting lower thus signals getting closer) while a positive value is a worsening.



(a) (b)

Figure 8.1: Mutual information based distance measure for excerpt DF3 (a) and M3 (b).

As we can see for the trained excerpts the difference between distances before and after the training is negative (thus the participants are closer to the conductor), while for the untrained ones is positive: this, even if it is just the introduction of a possible analysis, it is in line with the previous conclusions.

Conclusions

Conclusions and further investigations

The aim of my thesis was to show some analysis techniques useful when one is dealing with expressive gestures, in this case movements that mirror how listeners perceive the music.

My work was focused on proving two main points:

- firstly that, following the model of musical communication (Leman, [1]), the listeners movements are similar to the conductor ones because they should be the expression of the same intention;
- secondly, that familiarity with the music increases the correlation.

As we recall from the analysis (especially from Chapters 6 and 7), listeners gestures (in particular speed) are almost a precise replica of the music amplitude: even after the training, their movements are guided by the intensity of the music and the conductor intention is quite lost. With this analysis we have seen that correlation between participant and conductor exists: therefore some information encoded by the conductor in his/her gestures and conveyed by the orchestra through the music, was correctly received and "re-encoded" in gestures by the listeners. This correlation is however generally lower than the one between participants and music amplitude: it did increase with the training but surprisingly this is not a general statement.

From the direct inspection in Chapters 6 we found that some features of the movements, derived from the characteristics of the trained excerpts, remain also in the gestures performed during the un-trained one: in this case training did not increase the correlation. This phenomenon on the speed was also found in the comparison between conductor and participants movements in the correlation and mutual information analysis (Chapter 7 and 8): in fact for some excerpts (the un-trained case) there was a reduction of the values in the post-test, sign that in some way training contributed to "distance" conductor and listeners.

A possible continuation to the research would be in my opinion to repeat the experiment by adding also the visual stimulation during the training: in other words to let participants look at the conductors while conducting, so that maybe some more information concerning the intention can be retrieved.

Another aspect to take into account is the choice of the musical excerpt: for them to be meaningful there should have been in my opinion a stronger difference between the two executions, given that I obtained the most interesting results with excerpt 3 that was the one with both amplitude and tempo differences.

More research can be done on this topic by focusing on the analysis point of view but I think that my work, even if it has only scratched the surface of the argument, can be a well-rounded overview and an interesting starting point.

Bibliography

[1] M. Leman, *Embodied music cognition and mediation technology.* The MIT Press, 2008.

[2] P.-J. Maes, M. Leman, C. Palmer, and M. M. Wanderley, "Action-based effects on music perception," *Frontiers in Psycology*, vol. 4, no. 1008.

[3] F. Desmet, L. Nijs, M. Demey, M. Lesaffre, J.-P. Martens, and M. Leman, "Assessing a Clarinet Player's Performer Gesture in Relation to Locally Intended Musical Targets," *Journal of New Music Research*, vol. 41, no. 1, pp. 31–48, 2012.

[4] D. Amelynck, *The analysis of bodily gestures in response to music.* PhD thesis, Ghent University , A.Y. 2013-2014.

[5] F. Desmet, *A statistical framework for embodied music cognition.* PhD thesis, Ghent University, 2011.

[6] J. Ramsay, G. Hooker, and S. Graves, *Functional data analysis with R and Matlab.* Springer, February 2009.

[7] P. Hall and M. Hosseini-Nasab, "On properties of functional components analysis," *Journal of the Royal Statistical Society. Series B*, vol. 68, no. 1, pp. 109–126, 2006.

[8] D. Huron, *Sweet anticipation. Music and the psycology of expectations.* The MIT press, 2006.

[9] A. Cont, *Modeling musical anticipation: From the time of music to the music of time.* PhD thesis, Université Pierre et Marie Curie, 2008.

[10] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual information based registration of medical images: a survey," *IEEE Transactions on medical imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[11] J. E. Hudson, "Signal processing using mutual information," *IEEE Signal processing magazine*, vol. 23, no. 6, pp. 50–54, November 2006.

[12] E. King and A. Gritten, *New Perspectives on Music and Gesture*. SEMPRE Studies in The Psychology of Music, Ashgate Publishing Limited, 2013.